THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

Faculty of Business, Economics and Law
School of Economics

Smoke gets in your shape: analysing causal impacts of smoking on body mass index in Indonesia

A Thesis (ECON7931) submitted to the School of Economics, The University of Queensland, in partial fulfilment of the requirements for the degree of Master of Health Economics.

by Adrianna Bella, S.E., Universitas Indonesia

Acknowledgement Page

In the name of Allah, the Most Gracious and the Most Merciful. All praises to Allah for the strengths and His blessing to allow me to successfully finish this thesis.

Special appreciation goes to my supervisor Prof. KK Tang for his supervision and constant support and the opportunity he gave me to undertake research as a master student. His constructive suggestions, invaluable guidance, and encouragement have deeply inspired me to pursue my future career in economic research. It was a great privilege to be able to work under his supervision. My grateful thanks are also extended to my second supervisor, Dr Temesgen Kifle, for his support and knowledge especially relating to economic methods. His willingness to give his time so generously has been very much appreciated.

My sincere gratitude to my dearest husband, Vizzi Alvi Fitrah Nasution, who has constantly supported me and prayed for my success in pursuing my master's degree, and to my lovely daughter, Azrina Azkadina Nasution, for her unconditional love and support.

Last but not least, I thank the Australian Government for the Australia Awards Scholarship (AAS) that has provided financial and material support across my whole study period in Master of Health Economics, The University of Queensland.

Declaration Statement

I declare that the work presented in this Thesis is, to the best of my knowledge and belief, original and my own work, except as acknowledged in the text, and that material has not been submitted, either in whole or in part, for a degree at this or any other university.

Adrianna Bella

November 12th, 2019

Abstract

There are still inconclusive hypotheses and mixed evidence on the relationship between smoking and body weight. This study attempted to evaluate such a correlation in Indonesia, one of the world's largest tobacco consumers, that is experiencing the upward trend of overweight and obesity. Using panel data from the Indonesian Family Life Survey (IFLS) 1993–2014, I analysed the causal impacts of current smoking status and the number of cigarettes consumed daily on body mass index (BMI) in Indonesian men and women by applying the fixed effects with instrumental variables (FEIV) method. I also examined the role of former-smoking behaviours on BMI by applying the fixed-effect method to partly remove the endogeneity of former smoking. After controlling for the endogeneity, the results indicated that current smoking status and smoking intensity did not have any significant causal impacts on BMI for both males and females. I also found that former smoking status, smoking intensity before giving up smoking, duration of smoking before quitting smoking, and length of time since giving up smoking did not have a significant association with BMI in men and women. The findings could imply that smoking should not be used as a method to control weight and that one should not be discouraged from quitting smoking due to weight concern. From a public perspective, there was no evidence that anti-smoking policies could result in the unwanted rise of obesity nor could they lead to a double benefit—a decrease in both smoking and the obesity rate.

Key words: smoking, body weight, BMI, Indonesia

Table of Contents

Lists of Figures

Lists of Tables

Acronyms and abbreviations

BMI	Body Mass Index
CSNS	Current-smoker and non-smoker
CSFS	Current-smoker and former smoker
CS	Current smokers
DALY	Disability Adjusted Life Years
FE	Fixed Effects
FEIV	Fixed Effects with Instrumental Variables
GDP	Gross Domestic Product
IFLS	Indonesian Family Life Survey
OLS	Ordinary Least Square

1 Introduction

1.1 Background

Over the past three decades, there has been a dramatic increase in the global prevalence of adult obesity by 27.5% from 1980 to 2013 and, in 2013, it was estimated that around 13% of the world's population were obese (Ng et al., 2014; WHO, 2014). The problem of overweight and obesity has received increasing attention worldwide as it is linked to various adverse health effects such as diabetes, hypertension, heart diseases, cancers, mental illnesses, and premature death (Costa-Font & Gil, 2005; Hruby et al., 2016; Tremmel, Gerdtham, Nilsson, & Saha, 2017). Moreover, the overnutrition problem may directly impact the economy by increasing global healthcare costs, which accounted for around 2.8% of global GDP in 2014, and by indirectly impeding economic growth by reducing wages and the employment rate of obese people (Cawley, 2015; Tremmel et al., 2017).

Even though obesity was perceived as a problem of developed countries, currently the rate of obesity is rapidly increasing in developing countries while the rate in advanced countries recently reached a plateau (Ng et al., 2014). Developing countries, including Indonesia, are now home to around 62% of the obese population. As seen in Figure 1.1, the percentage of overweight[1] or obese people in Indonesia has doubled over the last two decades from about 15% in 1993 to approximately 32% in 2014 (Aizawa & Helble, 2017), which is equivalent to more than a 75% increase in the number of overweight and obese people (Helble & Francisco, 2017). The problem was responsible for 759,576 disability adjusted life years (DALYs) and 756,612 total productive years lost which may exacerbate the country's economy by incurring direct and indirect healthcare costs of 0.05% and 0.3%, respectively, of the country's gross domestic product (GDP) (Hruby et al., 2016). Despite the health and economic consequences of the rising overweight and obesity prevalence in Indonesia, there was still minimum concern by the government regarding the issue, proven by the absence of operational policy or an action plan to reduce this problem (WHO, 2016).

[1] The 'overweight' indicator in the graph includes people with overweight *or* obesity status, while the 'obesity' indicator only includes people with obesity status.

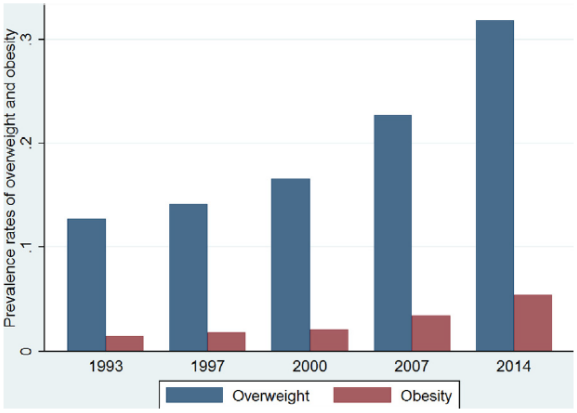

Figure 1 Overweight and obesity prevalence in Indonesia, 1993–2014

Source: Aizawa & Helble (2017)

On the other hand, smoking is currently still one of the primary sources of health problems that is responsible for a number of chronic diseases and smoking-attributable healthcare costs in Indonesia. Despite the global smoking prevalence dropping by more than seven percentage points since 2000, Indonesia has experienced an upsurge in the smoking rate, with adult smoking prevalence rising by 5.8% from 1995 to 2016 (World Bank, 2019; World Bank Group, 2018). Notably, male smoking prevalence jumped from 53.4% in 1995 to 68.1% in 2016, while female smoking prevalence jumped by 57.7% from 1990 to 2015 (Barber, Adioetomo, Ahsan, & Setyonaluri, 2008; Reitsma et al., 2017; World Bank Group, 2018). As a result, Indonesia has been acknowledged as the world's third largest home for smokers, with around 85 million of its population smoking (36.3% of the total population) and the country with the world's highest male smoking rate—68.1% (Reitsma et al., 2017; World Bank Group, 2018). In Indonesia, the smoking problem is expected to be responsible for more than one-fifth of all chronic diseases and more than 14% of all deaths (WHO Regional Office for South-East Asia, 2018; World Bank Group, 2018). The adverse health impacts of smoking were estimated to cost the country more than one billion US dollars per annum or around 8% of total public health expenditures (Barber et al., 2008; Tandon et al., 2016). Notwithstanding the high prevalence of smoking and its consequences for health and the economy, Indonesia

obtained deficient scores in WHO's MPOWER[2] measures and was the only country in South-East Asia which has not ratified the Framework Convention on Tobacco Control (FCTC) (Lian & Dorotheo, 2018; Reitsma et al., 2017).

Currently, Indonesia is experiencing an increasing overweight/obesity problem and the severe smoking issue remains; both have similar detrimental effects on population health and the economy. Therefore, understanding the causal relationship between smoking and obesity in this country is essential to completely comprehend the impact of public policies targeting one matter on the other matter. By way of illustration, if smoking is negatively related with body weight, then the implementation of anti-smoking campaigns should be coupled with the application of obesity-reducing policies. On the other hand, if smoking is positively associated with body weight, any tobacco-control policies should be massively encouraged due to its double return on smoking and obesity. Indeed, evaluating the causal effects of smoking on body weight in Indonesia may help identify the costs and benefits of any policy actions related to the two unhealthy behaviours.

Nevertheless, determining the direction of the association between smoking and body weight could be challenging since various theories underly the link between them. To date, the causal mechanism from smoking to body weight has been the subject of intense debate in the health literature. Some may argue that current smoking could decrease body weight by escalating metabolic rate and suppressing appetite (Audrain-Mcgovern & Benowitz, 2011; H. Chen, Saad, Sandow, & Bertrand, 2012; Chiolero, Faeh, Paccaud, & Cornuz, 2008), while others could also oppose the idea based on the fact that smoking may also reduce one's exercise capacity and induce fat accumulation as well as insulin resistance (Chiolero et al., 2008; Kvaavik, Meyer, & Tverdal, 2004; Rodriguez et al., 2010). Nonetheless, it may also be possible that both contrasting effects of smoking outweigh each other, so that smoking may not have any significant measurable impact on body weight. The various mechanisms linking smoking to body weight have attracted the attention of health economists, leading to evaluations of the correlation between smoking (or the

[2] MPOWER is six cost-effective indicators that may reduce tobacco consumption by WHO: 1) monitoring tobacco use and prevention policies, 2) protecting people from tobacco smoke, 3) offering help to quit tobacco use, 4) warning about the dangers of tobacco, 5) enforcing bans on tobacco advertising, promotion and sponsorship, and 6) raising taxes on tobacco (WHO, 2008).

costs of cigarettes) and body weight. The presence of endogeneity may even bring about more complications to determining the *causal* inference of smoking which may be sensitive to model specifications (see Chapter 2.2). As a result, there has also been inconclusive inference among economic publications regarding the causal or non-causal relationship between the two unhealthy behaviours: even though most papers found negative (positive) effects of smoking (quitting smoking) on body weight, some other research revealed positive (adverse) or insignificant effects of smoking (former smoking) on body weight.

Based on various underlying hypotheses and mixed evidence in the literature, there may not be a single generalisable impact of smoking on body weight. Moreover, the direction, significance, and magnitude of the effects of smoking may vary from one population to another as each society may have different socioeconomic circumstances and policies affecting both smoking and eating behaviours (Sohn, 2015). Therefore, this study set out to investigate the *causal* impacts of current smoking status and smoking intensity on body mass index (BMI) and the association of former-smoking indicators with BMI in Indonesia by using 21-year panel data collected in five waves from the Indonesian Family Life Survey (IFLS).

The present thesis may contribute to the literature in several ways. First, while most studies were conducted in developed countries and in China, I focused on the case of a developing country, Indonesia. Even though Indonesia has been one of the largest homes for smokers in the world and also has a growing obesity problem, to date, there seems to be neglect of this issue in the literature. Second, as many studies evaluated the role of cigarette costs on obesity, this study employed direct measures of smoking as the main predictors of body weight. Third, given that some earlier studies excluded female samples and combined people with different smoking status as a single reference group for current smokers, in this thesis, I separated the analysis by sex and by a comparison group of current smokers (non-smokers and former smokers). Fourth, I not only analysed the influence of current smoking, but I also evaluated the role of former smoking on body weight. Fifth, while other researchers provided only statistical associations between smoking and body weight, I aimed to examine *the causal* relationship between current (former) smoking and body weight by completely (partly) removing the endogeneity of smoking through fixed-effect instrumental variable (fixed-effect) method. In sum, my research appears to be the

first study analysing the causal effects of current smoking by entirely removing endogeneity problems, and is a pioneering study examining the link between former smoking and body weight in Indonesia.

The remaining part of this thesis proceeds as follows. In Chapter 2, I begin by laying out the theoretical dimensions of the research, unfurling the potential sources of endogeneity in the estimations, and looking at how previous studies determined the causal and non-causal relationship between smoking and body weight. The third and fourth chapters are concerned with the data and variables and the research methods used for this study, respectively. Chapter five presents the findings of the research as well as results of the sensitivity analysis. The sixth chapter of this paper provides a discussion of the findings, while the seventh chapter provides the conclusions of the overall study, including policy implications, limitations of the study, and suggestions for further research.

1.2 Research questions and objectives

The inconclusive findings in the literature regarding the effects of smoking on body weight and the current crucial problems of smoking and obesity in Indonesia encouraged me to examine the relationship between smoking and body weight in this country. Therefore, in this thesis I aimed to examine the *causal* correlation between current smoking on body weight and the association of former smoking with body weight in Indonesia. The primary outcome variable of interest as a proxy of body weight was BMI in kg/m^2. Two indicators measured current smoking: 1) dummy of current smoking status and 2) current smoking intensity as the average number of cigarettes consumed per day. There were four former-smoking behaviours employed in the study: a) former smoking status, b) the number of daily cigarettes consumed before quitting smoking, 3) duration of smoking before quitting smoking, and 4) length of time since giving up smoking. Due to the gap in the literature, I divided the analysis separately for males and females, rather than dropping female observations.

Additionally, as some previous studies mixed up the comparison group of current smokers, I separated the reference group into three categories: 1) non-current smokers which included former smokers and non-smokers, 2) former smokers who

had quit smoking, and 3) non-smokers who had never smoked previously. The effect of smoking intensity was also evaluated among current smokers only. Based on the explanation thus far, the main research questions can be recapitulated as follows:

1. Does being a current smoker have *causal* impacts on BMI in males and females compared with non-current smokers, non-smokers, and former smokers?
2. Does the average number of cigarettes consumed per day (smoking intensity) have *causal* effects on BMI in males and females compared with non-smokers, former smokers, and other current smokers?
3. Compared with non-smokers, how does being a former smoker, smoking intensity before quitting smoking, duration of smoking before giving up smoking, and length of time since quitting smoking affect BMI in males and females?

I also posed several additional research questions besides the main ones. The supplemental research questions were mostly based on the modification or the presence of other covariates. The other research questions are:

1. Does the addition of per capita oil and sugar consumption modify the impact of current smoking status and smoking intensity on BMI?
2. How do individual characteristics like age, age squared, marital status, working hours, and years of schooling influence BMI?
3. How do household characteristics—such as location in urban area and in Java province, per capita income, per capita oil consumption, and per capita sugar consumption—affect BMI?

The primary objectives of this study were to fill the gap in the literature by answering those research questions. However, establishing the heterogeneous impacts of smoking and exploring the link from smoking to body weight is beyond the scope of this study. This study may still leave some remaining questions about the effects of current smoking, smoking intensity, and former smoking on various quantiles of BMI and different weight status (underweight, healthy weight, overweight, and obesity) as well as the possible mechanisms of how the smoking-related variables may affect BMI in Indonesia.

2 Literature review

2.1 Theoretical foundation: the causal link between smoking and body weight/BMI

Previous studies in health literature have attempted to identify the underlying mechanisms of the causal effects of smoking on body weight or BMI. However, there is still no consensus in the literature regarding the issue. On the one hand, a considerable amount of literature has revealed the harmful effects of smoking through several potential channels. First, nicotine found in cigarettes may have a decreasing effect on body weight due to its positive impact on the metabolic rate which may then increase energy expenditure (Audrain-Mcgovern & Benowitz, 2011; Chiolero et al., 2008). Provided that obesity is the result of excess energy 'in' rather than energy 'out', the increase in energy 'out' by smoking may reduce one's weight or BMI. Second, smoking may have appetite-suppressant impacts on one's brain, so that smokers' appetite may decrease while their satiety may increase (Audrain-Mcgovern & Benowitz, 2011; H. Chen et al., 2012). The manipulated nervous system may, in turn, result in reduced caloric intake and a lower energy balance among smokers (Perkins et al., 1991). The theory also predicts that those who give up smoking (who decrease smoking intensity) may experience weight gain due to the absence (reduction) of appetite-suppressing effects from smoking. Taken together, these studies support the hypothesis that smoking may reduce body weight or BMI and that reducing or giving up smoking may exacerbate weight gain.

On the other hand, other studies analysed possible *positive* effects of smoking on body weight or BMI. First, smoking tends to decrease lung capacity which may result in lower physical activity (Kvaavik et al., 2004; Rodriguez et al., 2010). It is also possible that smoking increases fat accumulation (Chiolero et al., 2008). Smokers are also associated with increased fat intake, lower fruit and vegetable consumption, and unhealthy diet—which are favourable for weight increase (Palaniappan, Starkey, O'Loughlin, & Gray-Donald, 2001). Additionally, smoking may also increase the onset of hyperinsulinemia and insulin resistance which are associated with a higher probability of overweight and obesity (Chiolero et al., 2008; Schindler et al., 2006). Based on this theory, those who quit smoking may gain higher exercise capacity and have a lower probability of weight problems due to

insulin abnormalities. Together these studies provide valuable insights into the possibility that smoking may cause higher weight and BMI; while quitting smoking may cause a reduction in BMI, leading to a healthy weight and normal BMI.

The occurrence of each underlying mechanism in one country may not be the same as another country given a different environment and ongoing policies related to smoking and body weight (Sohn, 2015). For example, the type and amount of caloric intake as a substitute for smoking may depend on economic conditions, such as poverty, and the presence of food security in the society (Tanumihardjo et al., 2007). Another illustration was that different countries might have dissimilar public provisions that encourage physical activities and exercise. These circumstances may affect the mechanism of how smoking affects body weight. It may also be possible that both theories apply within individuals: insignificant effects of smoking on body weight may be because the positive impacts outweigh the negative ones. Therefore, one cannot determine which theory may be more relevant in which society. These different mechanisms and possible modifying environments may create unpredictable impacts of smoking on body weight measures, thus analysing the relationship in different countries may provide interesting results.

2.2 Endogeneity of smoking

2.2.1 Endogeneity problems: current smoking

The estimation of the impact of current smoking on BMI may not generate *causal* effects due to endogeneity issue. As suggested by Wooldridge (2013), the problem of endogeneity may arise due to three factors: 1) omitted variable bias (OVB), 2) reverse causality, and 3) measurement error. The omitted variable bias can be triggered by the presence of unobserved heterogeneity or time-invariant indicator and time-variant unobservable factors. The academic literature suggested that health perception and mental health are some of the predictors of both smoking and body weight which tend to vary over time (time-variant unobservable factors). Poor perception of health was associated with an increased probability of being a current smoker and a higher risk of obesity (Mendoza-Romero, Urbina, Cristancho-Montenegro, & Rombaldi, 2019). A growing body of literature has also investigated that mental health is one of the predictors of current smoking and obesity. One

study revealed that post-traumatic stress disorder might increase the risk of obesity by inducing unhealthy eating behaviours and raising the probability of smoking (Berk-Clark et al., 2018). Another study also observed that people with depression are more likely to be obese due to the increase in emotional eating and reduced exercise (Blaine, 2008). Depressed people also tend to become smokers as a way to obtain a 'self-medication benefit' (Weinberger et al., 2017). The studies presented thus far provide evidence that endogeneity—when estimating the impact of smoking on body weight—may be caused by omitted variable bias from unobserved time-variant factors, such as health perception and mental health.

Omitted variable from unobserved heterogeneity may also be responsible for the endogeneity of smoking. Adverse childhood experience (ACE) was one of the unobserved time-invariant factors of both smoking and body weight. Notably, several studies determined that people experiencing physical abuse, sexual abuse, verbal abuse, familial substance abuse, witnessing domestic violence, parental separation, household members' imprisonment, and the presence of mentally ill person as a household member in childhood may have higher likelihood to be smokers and become obese compared with those without adverse childhood experiences (Davis, Barnes, Gross, Ryder, & Shlafer, 2019; Ford et al., 2011; Rehkopf et al., 2016). Additionally, it was also possible that the decision to smoke and the diet-related behaviours may be influenced by one's time preference (Robb, Huston, & Finke, 2008). People with relatively higher rates of time preference and higher discount rate of future state tend to put more value on present utility over the future so that they are more likely to enjoy smoking and unhealthy diet in the present time and less likely to be concerned about their health in the future. In all the studies reviewed here, unobserved heterogeneity was recognised as one of the sources of endogeneity in current smoking.

Endogeneity of current smoking may also originate from reverse causality between smoking and body weight. Even though a large number of studies have explored possible mechanisms of either positive or negative effects of smoking on body weight, it is also possible that body weight affects one's decision to smoke. As an illustration, Cawley, Markowitz, & Tauras (2004) and Kim (2018) found that people with overweight status as well as people with a perception of being overweight/obese were more likely to smoke as a means of losing weight, while

Carreras-Torres et al. (2018) established that each additional standard deviation in BMI increases the risk of being a smoker. The measurement error of smoking may also occur since smoking and the average number of cigarettes smoked per day were self-reported by respondents. Overall, the evidence reviewed here seems to suggest a pertinent role in the existence of endogeneity in current smoking due to omitted variable bias, reverse causality, and measurement errors.

The author of a very similar study in Indonesia argued that endogeneity problems in the relationship between smoking and body weight might only come from unobserved heterogeneity, while reverse causality and time-variant errors were unlikely to be the source of endogeneity (Sohn, 2015). The first argument was that reverse causality might not be the case in Indonesia because of two main reasons: 1) little evidence regarding the use of smoking on weight control in Indonesia and 2) the prevalence of women smoking as a way to control weight is very low in Indonesia. Another argument was the assumption that unobserved time-variant factors was minimal by comparing the magnitude difference between the OLS and fixed effects (FE) results. Sohn also assumed that unobserved time-variant factors may only originate from individuals' concern about health over time. I disagree with the statement by Sohn (2015) and contend that reverse causality and unobserved time-variant factors may still occur in Indonesia. First, one cannot simply ignore the fact that reverse causality did not take place only based on the absence of evidence. Given that the volume of Indonesia's research publications are still low compared with advanced countries (Indonesia National Research and Inovation Agency, 2016), the unavailability of the evidence may not reflect the absence of the problem. Additionally, reverse causality may not only happen in women, as Sohn suggested, since Carreras-Torres et al. (2018) found that the higher likelihood of being smokers as body weight increased affected both males and females. Even if reverse causality happened only in women in Indonesia, the low prevalence of women who smoke in Indonesia might not be used as the reason for ignoring the issue. Second, as presented in this study, previously unobserved time-variant factors may not only come from health concern changes over time but may also come from the emergence of mental health conditions that may induce both overeating and smoking behaviours. In view of all that has been mentioned so far, the problem of endogeneity in the relationship between smoking and body weight in Indonesia may originate from unobserved

heterogeneity, time-variant factors, and reverse causality. Therefore, removing only the partial endogeneity problem may still generate bias outcomes.

2.2.2 Endogeneity problems: former smoking

There was limited evidence on the endogeneity of former smoking, but several studies provided arguments for former smoking also being endogenous (Courtemanche, Tchernis, & Ukert, 2018; Kasteridis & Yen, 2012; Pieroni & Salmasi, 2016). Generally, the sources of endogeneity problems of quitting smoking may be similar to the one of current smoking which may come from unobserved heterogeneity, time-variant unobservable factors, and reverse causality. Au, Hauck, & Hollingsworth (2013) pointed out that unobserved heterogeneity influenced the decision to quit smoking and body weight among males and females, but they could not identify the factors in the study. There seemed to remain time-invariant factors influencing both former smoking and body weight about which relatively little is known.

Time-variant factors may also be the case of the endogeneity of former smoking. Another study signified that self-assessed health status of smokers, which tended to change over time, could affect smokers' decision to give up smoking (Amir, 1996). Healthy smokers may quit smoking as a way to prevent future illnesses, while unhealthy smokers may give up smoking as a healing method. At the same time, a low perception on one's health may also elevate the risk of obesity (Mendoza-Romero et al., 2019). The absence or recovery from depression was expected to ease and induce the commitment to quit smoking (Reid & Ledgerwood, 2015) as well as to promote healthy-eating behaviours (Blaine, 2008). Reverse causality may also be why obese smokers were more reluctant to quit smoking due to their concern for weight gain after smoking cessation (Bush, Hsu, Levine, Magnusson, & Miles, 2014). Collectively, these ideas outline the existence of endogeneity in former smoking behaviours.

2.3 Previous studies: the effects of smoking on body weight measures

Analysis of unhealthy behaviours, such as smoking, drinking, drug use, and obesity, have been attracting health economists since the 1970s and comprise at least half of the sub-specialty subjects of health economics in the early 2000s (Cawley & Ruhm, 2011). Examination of the causes and consequences of unhealthy behaviours is necessary because they lead to at least 48% of all deaths (Cawley & Ruhm, 2011). While previous health economic researchers have examined the link between unhealthy behaviours and economic indicators like income, time preference, price elasticities, and education, a growing body of literature reveals analysis on the relationship between two unhealthy behaviours, such as smoking and drinking, obesity and smoking, smoking and drug use, drinking and obesity, and risky sex and drinking (Courtemanche et al., 2018). Comprehension of the correlation between two unhealthy behaviours is necessary to fully understand the consequences of one unhealthy behaviour on another.

Nevertheless, the analysis of the *causal* effects of health behaviours on economic outcomes or other health behaviours may face considerable challenges as there usually exists reverse causation from the outcome indicators to the health behaviours and confounding factors which influence both outcomes and health behaviours, which are usually referred to as the endogeneity problems in economics (Cawley & Ruhm, 2011). Ideally, the analysis of health behaviours should be conducted by implementing randomised control trials to isolate the impacts of health behaviours on specific outcomes; however, such methods mostly are unethical or not feasible to implement (Cawley & Ruhm, 2011; Webb & Bain, 2011). As a result, the approach of health economics to fully remove the problem of endogeneity in health behaviours is vital to draw causal effects of the behaviours on economic outcomes or other health behaviours.

This sub-section provides a review of the health economics literature which has tried to examine the causal and non-causal relationship between smoking and body weight (obesity). I also included a brief description of other studies in the health literature that analysed the non-causal inference in similar issues to provide insights which may not be uncovered in the economic literature. In this sub-section, I will

first review the initial debate on the link between cigarette costs and obesity, and then move into the literature examining the role of each smoking-related variable used in this study.

2.3.1 The role of cigarette costs on the rise of obesity

The interest of health economists in exploring the association between smoking and body weight was triggered by the dramatic increase in the obesity rate in the United States since the 1970s (Chou, Grossman, & Saffer, 2004; Wehby & Courtemanche, 2012). At the same time, the prevalence of smoking had dropped significantly while the price of cigarettes has risen since the 1960s. In the early 2000s, previous studies in the field of economics started investigating the role of smoking on body weight by exploring the relationship between cigarette costs, such as cigarette price and tax, and body weight or BMI. The idea was that the increase in cigarette costs might reduce the affordability of cigarettes as well as encourage smokers to quit or reduce smoking and substitute smoking with higher caloric intake.

The first study in this area was conducted by Chou et al. (2004) and Rashad & Grossman (2004). Chou et al. (2004) explored the causal effect of cigarette prices on BMI and the probability of being obese in the United States by using panel data and fixed-effect models. The results showed that cigarette prices were positively correlated with BMI and the likelihood of being obese. They concluded that the obesity rise in the United States was an 'unintended consequence of the anti-smoking campaign'. In the same year of 2004, Rashad & Grossman (2004) also supported the study by Chou et al. (2004) by finding that every 10% rise in real cigarette price may increase the prevalence of obesity in the United States by 2%, and that cigarette price was responsible for around 20% of the rise of obesity from the 1960s to 2000. They also proved the increase in the obesity rate in the United States was an unanticipated consequence of the rise in cigarette taxes and anti-smoking programs. The results from the two mentioned studies dissatisfied Gruber & Frakes (2006) and triggered them to revisit the impact of cigarette cost on BMI and obesity by re-running the data used by Chou et al. (2004). Gruber & Frakes modified several things from the study by Chou et al.—such as excluding the elderly, employing cigarette taxes instead of cigarette price, which may suffer endogeneity problems, and the inclusion of time dummies—and found that cigarette price is *negatively*

correlated with both BMI and body weight and that Chou et al. provided 'wrong-signed' findings. The results implied that as cigarette tax increased and smoking or smoking intensity decreased, BMI and the probability of being obese may decrease. Particularly, every US$1 increase in cigarette tax may reduce the obesity rate by 0.02 percentage point.

The debate in the economic literature regarding the impact of smoking on body weight continued as Baum (2009) provided another research finding comparing the methods and results by both Chou et al. (2004) and Gruber & Frakes (2006). He attempted to resolve the disagreement in the literature by finding that either cigarette price or tax may *increase* BMI as well as the probability of being overweight and obese. His findings opposed the ones established by Gruber & Frakes (2006) while they supported the results from two preliminary studies by Chou et al. (2004) and Rashad & Grossman (2004) that the increase in cigarette cost, which may decrease smoking intensity and prevalence in the society, may increase various measures of body weight. Nonnemaker, Finkelstein, Engelen, Hoerger, & Farrelly (2009) also did similar research to Baum's (2009) by evaluating prior studies by Chou et al. (2004) and Gruber & Frakes (2006). Their research was triggered by the statements from Gruber & Frakes (2006) and Chou et al. (2004) that their results provided an exaggerated effect of cigarette cost on body weight which was non-supported in the medical literature. Nonnemaker et al. (2009) found that the overall effects of cigarette price on body weight were insignificant, while there was only a slightly positive impact on former smokers. Taken together, they established that there was *not enough evidence* to support the claim that the change in the obesity rate in the United States from the 1970s to 2000s was due to the increase in cigarette cost.

Other studies continued to evaluate the conflict of previous studies, analysing the effect of the cost of cigarettes on body weight measures by internalising model specifications that failed to be observed in previous studies. Courtemanche (2009) disagreed with all the studies mentioned above and argued that the effects of cigarette cost on the obesity rate might provide invalid results in the short-run due to misspecification of models (e.g. ignoring delayed impacts of cigarette cost on body weight). Based on the idea that the impact of smoking cessation on weight gain was likely to be temporary, he included lagged cigarette price and tax to see their long-

term impacts on body weight measures. The coefficient signs of the finding supported the one by Gruber & Frakes (2006) that higher cigarette price or tax may *reduce* BMI and overweight/obesity status in the long run, particularly a one dollar increase in cigarette tax or price may decrease BMI by 0.13-0.59 points and the prevalence of obesity by 1.1-3.6 percentage points. Courtemanche (2009) also provides the mechanism of the positive effects of cigarette cost on body weight by finding that the increase in cigarette cost may increase exercise, decrease fat intake, and increase vegetable intake. Based on the modified models by Courtemanche (2009), Wehby & Courtemanche (2012) pointed out that the long-term effects of cigarette cost may not be generalised to other populations and that there may be heterogeneous effects over different tails of BMI and among different groups of people. Their overall findings were that cigarette price *reduced* BMI and that the effect was more substantial as people moved from lower to the higher tail of BMI (the effect of cigarette price was more tremendous as BMI was higher). The heterogeneous negative influence of cigarette price was also observed across different groups in which the effects of cigarette price on BMI was larger for those who were black (compared with those who were white); highly educated (compared with those with high school education or less); younger, aged less than 60 years (compared with those aged 60 years or more); and female (compared with their male counterparts).

The evidence presented in this section suggests that health economists have paid attention to the potential relationship between the rise of the obesity rate and the drop of cigarette price or tax over decades, particularly in the United States. In view of all that has been mentioned so far, there seemed to be no consensus regarding the impact of cigarette cost on body weight. Some research found obesity as the unintended consequence of the increase in cigarette cost (Baum, 2009; Chou et al., 2004; Rashad & Grossman, 2004), others established negative effects of cigarette cost on body weight (Courtemanche, 2009; Gruber & Frakes, 2006; Wehby & Courtemanche, 2012), while one study observed no significant impacts of cigarette cost on body weight measures (Nonnemaker et al., 2009). Since different medical and biological theories underlined every finding on the mechanism from smoking to body weight, one cannot decide the correct direction of the effects of cigarette cost on obesity and body weight. Nonetheless, all studies presented in this section used indirect indicators of smoking (cigarette price and tax) as the main predictors of

body weight which may not reflect the real effects of smoking behaviour on body weight measures. Moreover, all studies mentioned above reflected the condition in an advanced country, the United States, so that the condition may be different in developing countries.

2.3.2 Direct measures of current smoking on body weight

More recent attention has focused on the consequences of direct smoking indicators, such as smoking status or smoking intensity, on various measures of body weight like weight, BMI, or weight status. Many of the current studies were conducted in various countries with different levels of income and were published in economic and health journals. Currently, there are still disagreements on the direction of the impacts of direct smoking indicators on body weight among the studies. The review of previous studies, analysing the effects of current smoking status and smoking intensity on body weight measures, are presented in the following paragraphs.

Current Smoking Status

In the economic literature, one of the notable differences among the papers was the use of methodologies and the way they addressed the problem of endogeneity of smoking-related variables. Among the studies analysing the effect of current smoking status with economic methods, half of them eradicate endogeneity issues while others still suffered from endogeneity bias. A study by Huffman & Rizov (2007) explored some determinants of body weight and BMI in Russia by implementing ordinary least squares (OLS) method and separating the sample into male and female. Their findings indicated that being a current smoker was associated with lower weight and BMI in full, men, and women samples, compared to those who did not currently smoke (both former and non-smokers). Mainly, being a smoker may reduce weight and BMI by 0.03–0.08 point. However, the impact of smoking on females was only significant at 10% level of significance. The OLS method was extended to a quantile regression and ordered probit by Plurphanswat & Rodu (2014) in order to examine whether the impacts of smoking may vary depending on the tail of BMI distribution and weight status in the United States. They also compared current smokers only

with those who never smoked by removing former smokers from the reference group and separating men and women into different regressions. The results showed that current smokers might have lower BMI by 1.5 to 2 points in women and men respectively. The association between smoking status and BMI was negative across all BMI distributions, but there was no observed increasing or decreasing trend of the magnitude of the impacts. Interestingly, Plurphanswat & Rodu (2014) found that being a current smoker may increase the probability of being underweight and overweight as well as having healthy weight in both men and women (the effects on overweight status was only significant in women). However, it may decrease the likelihood of being obese for both males and females. These two studies may provide some insights into the economic literature that reports smoking may be correlated with body weight indicators and that the effects may be heterogeneous across different BMI quantile or weight status. Nevertheless, the results seemed to be specious as there may still be endogeneity problems which lead to bias inference.

Two other economic studies tried to estimate causal effects of current smoking status on body weight by removing endogeneity issues. A study by Wang (2015) in China was similar to the models established by Plurphanswat & Rodu (2014) that were used to analyse the effects of smoking status on four weight categories: underweight, healthy weight, overweight, and obese. The distinctive feature of this study was the implementation of the two-stage residual inclusion (2SRI) method with pooled OLS (panel data), which was equivalent with the two-stage least square (2SLS) method for regressions with the binomial or multinomial dependent variable, by employing local tobacco production as an instrument for smoking. Having controlled the endogeneity issues, the author observed that current smoking status increased the probability of being underweight by 0.9 percentage points (pp) and having healthy weight by 5.3 pp while it decreased the likelihood of being overweight and obese by 6.5 pp and 5.1 pp respectively. Despite the sophisticated modelling, this paper did not separate the comparison groups for smokers into former and non-smokers and ignored the possibility of different smoking effects on women and men. In the same publication year, Sohn (2015) estimated the effects of smoking on body weight in Indonesia by using panel data similar to what I am using from the Indonesian Family Life Survey (IFLS). Unlike Wang, Sohn (2015) employed a fixed-effect model so that he only removed endogeneity problems from unobserved

heterogeneity. His findings from fixed-effect models showed negative effects of current smoking on body weight measures, whereby a current smoker may decrease weight and BMI by 0.92–0.98 kg and 0.35–0.37 kg/m^2, respectively. The effect of smoking status on the probability of being obese was insignificant before adding income covariates and was only significant at 10% after including earning variables. Even though Sohn's study generated similar direction of the effects of smoking on body weight with the study authored by Wang (2015), the study by Sohn (2015) still suffered from endogeneity coming from reverse causality and time-variant errors. Moreover, this study dropped the female sample, assuming that 'smoking is an exclusively male habit in Indonesia', and combined both former smokers and non-smokers as a comparison group of those currently smoking.

In the field of public health and other health sectors, the observed direction of the effects of smoking on body weight tended to vary from one study to another: some research observed the negative influence of smoking status, while others found positive or insignificant effects of the indicator. However, in most of these studies, researchers analysed cross-sectional association, thus no causal effects were obtained. A study by Flegal (2007) estimated the relationship between the change in smoking prevalence and obesity prevalence in the United States. The findings suggested that even though current smokers tend to have a lower probability of obesity compared with non-current-smokers, the overall prevalence of smoking only had small impacts on the prevalence of obesity in the United States. This study's finding was contrary to findings from previous studies examining the impacts of cigarette cost on body weight in the United States. Four cross-sectional studies in China (Xu, Yin, & Wang, 2007), Japan (Watanabe et al., 2016), the United Kingdom (Dare, Mackay, & Pell, 2015), and the United States (Kaufman, Augustson, & Patrick, 2012) found that current smoking status was negatively associated with body weight compared with non-smoking status. However, the study in Japan couldn't find significant effects of current smoking status in women; the one in China didn't include females in the sample, thus no analysis conducted on women; the research in the United Kingdom found significant effects for both women and men but no significant effects among people aged 40 years and below; and the study in the United States didn't divide male and female samples but found that sedentary behaviours moderated the effect of smoking status on body weight. One of the critical attributes of these studies was

the separation of non-smokers and former smokers as the reference group of smoking. However, one study found conflicting results that after approximately 5.5 years of follow-up those who currently smoked at baseline were found to gain more weight compared with non-smokers (Guerra, Stringhini, Vollenweider, Waeber, & Marques-Vidal, 2015). Although the author did not provide the mechanism underlying the result, it could be merely due to smokers quitting during follow-up or possibly because smokers tend to have unhealthy behaviours promoting weight gain, or it could be the impacts of smoking on higher fat accumulation and waist circumference. Overall, even though most cross-sectional studies from the health sector found adverse effects of current smoking status on body weight, it is still possible that current smokers have higher weight or experienced relatively more substantial weight gain than non-smokers.

The intensity of current smoking

There are relatively more studies analysing the effects of the average number of cigarettes consumed in body measures than those examining current smoking status in both the economic- and health-related literature. One probable reason is that smoking status cannot provide an estimate of the effects of smoking addiction by treating all smokers as similar (e.g. treating equivalently light smokers, heavy smokers, social smokers, and occasional smokers) (Z. Chen, Yen, & Eastwood, 2007). Therefore, other studies preferred to use the number of cigarettes smoked as an indicator of smoking. Currently, there is still disagreement among the studies on how to assess the effects of smoking on body weight measures. Some studies found that the number of cigarettes consumed may positively affect lower body weight status and negatively affect the upper tail of body weight measures, others found constant negative impacts at any status of body weight, while a few others found constant positive or non-significant effects of smoking intensity.

Among the six economic studies reviewed, all the authors fully or partially removed endogeneity problems: in four papers, the authors applied two-stage least square (2SLS) (Courtemanche et al., 2018; Fang, Ali, & Rizzo, 2009; Wang, 2015; Wehby, Murray, Wilcox, & Lie, 2012), in one paper the simultaneous equation model was implemented (Z. Chen et al., 2007), and in one study FE were used (Sohn, 2015). In his study, Sohn (2015) analysed the role of smoking intensity by using FE

and by partially removing the endogeneity of smoking. He found that consuming an average of 1 to 19 cigarettes per day was associated with a reduction in weight, BMI, and the probability of being obese by around 0.4–0.7, 0.2–0.3, and 2.4–2.7 percentage points, respectively. However, he observed that consuming more than 20 cigarettes a day did not have any significant influence on all body weight measures and that the effects of smoking intensity on body weight and BMI were relatively small.

In four other studies where the authors used instrumental variables to tackle endogeneity problems, various heterogeneous impacts of smoking on body weight measures were found. The causal effects of smoking intensity on body measures was, to my knowledge, first studied by Fang et al. (2009) in a sample of males in China by using the community average number of cigarettes smoked per day and the community price per pack of the most commonly sold cigarette brand as instruments of smoking intensity. They obtained four measures of body weight in their study: 1) BMI, 2) BMI distribution (quantile), 3) four BMI categories, and 4) overweight (including obesity) status. Fang et al. employed four models of two-stage least squares (2SLS), quantile regression with two-stage residual inclusion (2SRI), ordered probit with 2SRI, and binomial probit with 2SRI to estimate four body measures respectively. Their results showed that the average number of cigarettes consumed per day was negatively correlated with BMI and the probability of being overweight/obese. Interestingly, smoking consistently had negative effects on the 25^{th}, 50^{th}, and 75^{th} quantile of BMI compared with the lowest quantile (no observed effects for the highest quantile). The rate of smoking addiction also had a positive impact on the probability of being underweight and having healthy weight, while it had a negative relationship with pre-obese and obesity status. Instead of sophisticated modelling, some other health economists argued that the study provided inconsistent estimates of quantile regression and suffered invalid instruments (Sohn, 2015; Wang, 2015; Wehby et al., 2012).

Wehby et al. (2012) tried to establish a better study than the one by Fang et al. by the implementation of a more convincing instrument from genetic indicators. Analysing the effects of smoking intensity in Norwegian mothers on body weight before pregnancy, this study found that the number of cigarettes smoked per day may increase BMI at the 0.1 and 0.25 quintile while reducing BMI at the 0.9 quintile. However, the study also showed contrasting results that the average daily

consumption of cigarettes pre-pregnancy may decrease the probability of being underweight and raise the likelihood of becoming overweight, but no observed significant effects on obesity status. Although the study used seemingly compelling instruments, this study was no different from the one by Fang et al. (2009) due to inconsistent results across different body weight specifications. Attempting to generate more plausible causal impacts of smoking intensity on body weight measures in China, Wang (2015) employed similar methods as Fang et al. (2009) with a different data set and instrumental variables for measuring smoking. In the same vein as Fang et al., he found that an extra cigarette smoked per day may elevate the probability of being underweight and normal weight but reduces the likelihood of being overweight and obese. Courtemanche et al. (2018) extended all the mentioned studies by analysing the short-term and long-term effects of cigarettes consumed on BMI and BMI quintiles. The outcomes of the study indicated that smoking intensity was inversely associated with BMI in both the short- and long-run with larger effects found in the long run. Unlike the studies by Fang et al. (2009) and Wehby et al. (2012), this study observed consistent adverse effects of average cigarette consumption across all BMI quintiles. Nontheless, the results by Wang (2015) and Courtemanche et al. (2018) may be subject to the pooling of men and women in the same regressions.

In contrast to earlier findings, Z. Chen et al. (2007) found insignificant long-run causal effects of smoking intensity on BMI in the United States by using a simultaneous equation model (SEM). Finding significant negative effects of smoking addiction on BMI under OLS regression, the authors of this study suggested that other studies, failing to remove endogeneity issues, may provide bias negative effects of smoking intensity. One possible explanation of the study outcomes was that the slimming effect of smoking might be temporary. A necessary inference in the study was that cigarette smoking should not be used as a method to control weight since the results indicated insignificant effects of cigarette quantity.

Taken together, the authors of the mentioned studies used economic methods to analyse the role of smoking intensity not only on single measures of body weight, like BMI or weight, but also on different BMI distributions and weight status. In some studies, the authors found negative causal effects of smoking intensity on BMI (Fang et al., 2009; Sohn, 2015), while one study revealed no significant causal

relationship between cigarette consumption and BMI (Z. Chen et al., 2007). Other studies observing the causal effects of the number of cigarettes smoked in four weight categories (underweight, normal weight, pre-obese, and obese) and BMI quintiles found various inconclusive effects (Courtemanche et al., 2018; Fang et al., 2009; Wang, 2015). Despite the implementation of strong models to assess causal impacts, the authors of most studies did not separate non-smokers and former smokers into two separate reference groups for current smokers, nor did they include the analysis of smoking intensity among current smokers only.

Evidence in health-related literature showed more variations in the relationship between smoking intensity and body weight measures. In most studies, the authors classified smoking intensity into several groups, such as light smokers, moderate smokers, and heavy smokers. The authors of some studies also conducted the analysis among current smokers only. In comparison with non-smokers who never smoked, the authors of one paper found that the number of cigarettes smoked per day was negatively correlated with BMI for light, moderate, and heavy smokers (Xu et al., 2007), while another study suggested a contrasting idea, namely, that the prevalence of obesity was more likely to increase as the number of cigarettes consumed per day increased (Watanabe et al., 2016). In two studies, the authors observed that light and moderate smokers tended to have a lower probability of overweight and obesity, while heavy smokers were more prone to be obese compared with non-smokers (Chiolero, Jacot-Sadowski, Faeh, Paccaud, & Cornuz, 2007; Dare et al., 2015). The relationship might be even more complicated as authors of other studies found that the number of cigarettes smoked daily had no significant link with BMI and the probability of being overweight/obese (Clair et al., 2011; U. John, Hanke, Rumpf, & Thyrian, 2005). With respect to the impact of smoking intensity among current smokers only, it was suggested that the number of cigarettes consumed per day may be positively correlated with BMI and the likelihood of being obese, while at the same time the same research also reported that current smokers on average have lower probability of being obese compared with non-smokers (Bamia, Trichopoulou, Lenas, & Trichopoulos, 2004; Dare et al., 2015). Together, the studies in health literature indicated the inconclusive effects of smoking intensity on BMI and obesity. All the studies reviewed so far, however, suffer from being cross-sectional or

population-based and from not dealing with endogeneity issues so that there was no observed causal impact of the rate of smoking intensity

Some other health studies supported the idea that smoking intensity may increase one's BMI and the probability of being overweight/obese (compared with non-smokers) by providing several reasons. First, current smokers with higher smoking intensity tended to have a poor quality diet, which may escalate the probability of becoming obese (Alkerwi et al., 2017). It was also estimated that smokers with higher smoking addiction were more likely to have higher fat intake and lower fruit and vegetable intake (Palaniappan et al., 2001). Second, the increase in the smoking intensity was linked with lower physical activities and a lower level of exercise (Chiolero et al., 2008). Third, the escalated rate of smoking addiction was associated with the increased risk of insulin resistance (Haj Mouhamed et al., 2013), which was expected to be one of the sources of increased body weight (Schindler et al., 2006). Although the existence of these supportive studies may strengthen the idea of the positive effects of cigarette consumption, it may not rule out the possibilities that smoking intensity may have opposite or non-significant impacts on body measures.

2.3.3 The role of former smoking

There were quite limited numbers of economic studies examining the role of former smoking on body weight measures, but there was abundant literature in health disciplines on this issue. In previous studies, the authors mostly analysed former smoking status as the primary variable of interest, but few authors also explored the effects of duration of smoking before quitting, smoking intensity before giving up smoking, and the period after quitting smoking. While the authors of studies regarding current smoking mostly analysed the impacts of smoking on BMI, the authors investigating former smoking measured the effects on weight gain. Earlier studies about the effects of former smoking on body weight measures in economics and health literature are evaluated in the following paragraphs.

Evidence from five studies exploring the role of former smoking behaviours using econometric methods indicated that two studies treated quitting smoking as exogenous (Z. Chen et al., 2007; Wang, 2015), while the authors of the other three studies assumed former smoking as endogenous and thus used specific methods to

deal with possible endogeneity problems (Courtemanche et al., 2018; Kasteridis & Yen, 2012; Pieroni & Salmasi, 2016). Z. Chen et al. (2007) estimated both current cigarette consumption and quitting smoking using OLS and simultaneous equation model (SEM) to account for the endogeneity of current cigarette consumption. However, the dummy variable, quitting smoking, was treated as exogenous and the inclusion of this variable meant the addition of more explanatory power to the regressions. The results indicated that former smokers might have higher BMI around 0.445 to 0.335 percentage points in both OLS and SEM, respectively compared with non-smokers. A study by Plurphanswat & Rodu (2014) estimated the impacts of former smoking not only on BMI, but also on BMI quintiles and weight categories. In comparison with non-smokers, they observed significant positive effects of being a former smoker on BMI only in women and not in men, and at only 10% level of significance. The former smoking status also had consistent positive effects on BMI across all BMI quintiles for women at 10% and 5% level of significance, while there was no significant effect for men in any percentile of BMI. Being a former smoker may decrease the probability of being underweight, overweight, and having normal weight but may increase the probability of being obese in women. Nonetheless, the results were mostly significant but only under 10% confidence level. These two studies generally found that quitting smoking may increase BMI compared with non-smokers.

The authors of three economic studies estimated the effects of former smoking on body weight by implementing certain methods to remove endogeneity problems. Pieroni & Salmasi (2016) attempted to examine the causal effects of quitting smoking on BMI by implementing a difference-in-difference model to remove endogeneity problems from unobserved heterogeneity. The results showed that individuals who quit smoking may have a higher BMI of around 0.5-0.6 unit compared with current smokers or non-smokers. The positive effects were also observed in 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of BMI compared with current smokers. Overall, subject to the limitations on generating the causal effects of quitting smoking, this study suggested that quitting smoking may have positive impacts on weight gain after smoking cessation. Courtemanche et al. (2018) also provided additional knowledge to the literature by examining the causal impacts of quitting smoking using instrumental variables. The instruments seem very strong since they used the intervention data

from a randomised control trial: the intervention was receiving an inhaled bronchodilator versus receiving an inhaled placebo. The intervention was meant to support smokers to quit smoking in four months. The findings showed that quitting smoking may increase BMI in both the short-run (1 year) and long-run (5 year), with tiny differences in magnitude. However, this study did not determine the comparison group of the quitters: whether non-smokers, current smokers, or combined non and current smokers. In another study, Kasteridis & Yen (2012) examined the causal relationship between duration after quitting smoking on BMI across three different groups by using cigarette tax as an instrumental variable of quitting smoking. They estimated that quitting smoking may increase BMI and that the positive effects became larger as the duration of smoking absence increased. The observed impacts were also larger among older people (age ≥ 41 years old) compared with younger people and greater in females compared with males. Overall, after eradicating endogeneity problems, the studies in the economic literature suggested that quitting smoking may contribute to higher BMI.

Previous studies from health disciplines mostly found similar results as the ones in the economic literature even without removing endogeneity problems: quitting smoking may be positively associated with BMI after cessation (Filozof, Fernández Pinilla, & Fernández-Cruz, 2004; Piirtola et al., 2018; Taleb et al., 2017; Veldheer, Yingst, Zhu, & Foulds, 2015; Watanabe et al., 2016). However, some research observed no significant effects of former smoking status on weight gain or BMI compared with non-smokers (Guerra et al., 2015; U. John et al., 2005; Køster-Rasmussen et al., 2015). A study also pointed out that even though former smokers may gain weight after smoking cessation, they would, in the end, have a similar weight as non-smokers (Køster-Rasmussen et al., 2015). Contrasting with the majority of studies in this issue, one group of researchers denoted that the concurrent emphasis of previous papers on higher BMI or weight gain after smoking cessation might contribute to the misleading deduction in the literature and society that higher BMI was one of the inevitable consequences of quitting smoking (Pisinger, Nielsen, Kuhlmann, & Rosthøj, 2017). The researchers emphasised that earlier research observing weight gain after smoking cessation may be due to a short follow-up period when weight gain may be the immediate impact of quitting smoking. Pisinger et al. found in a 5-year cohort study that around 13% and 4% of ex-smokers,

respectively, experienced reduced or maintained weight. Former smokers with lighter smoking intensity and overweight/obesity status may obtain higher weight reduction compared with former heavy smokers and those with normal weight status. Overall, even though most previous health literature suggested higher BMI of former smokers compared with non-smokers, it was also possible that quitting smoking may have insignificant or negative impacts on body weight.

2.3.4 Summary of the literatures and the study gap

There have been some contrasting hypotheses indicating that smoking may not only negatively affect body weight but may also contribute to an increase in body weight. The theories also implied that quitting smoking may increase body weight after the slimming effects of smoking, or it may decrease body weight by removing the weight-gain effect of smoking. Economic and health-related literature responded to the phenomenon by also finding inconclusive effects of cigarette cost, current smoking status, cigarette consumption, and quitting smoking on BMI, BMI quintiles, weight, and weight status: the former indicators may affect the latter positively, negatively, or insignificantly.

Several previous kinds of literature possessed a number of limitations. First, while there are already a few papers indicating causal impacts of smoking on body weight measures, some other studies still suffered from the endogeneity of smoking and thus provided bias results. Moreover, the authors of most studies analysed data from developed countries, or from China. Only one study was found to examine the relationship between smoking and body weight in Indonesia—a developing country. The authors of some papers also excluded female samples due to the low number of female smokers or pooled men and women together in one sample. Given the different nature of body weight and BMI of males and females (Ritchie & Roser, 2019), pooling them together may provide misleading results. Additionally, the low number of smokers among women may not be a valid reason for dropping women from the sample. Although Watanabe et al. (2016) found diverse effects from the different comparison groups, authors of several studies also compared current smokers with both former and non-smokers without separating them as two reference groups. Regarding the analysis of smoking intensity, some researchers examined the impact of smoking intensity among current smokers only; however, most studies did not

include the analysis. Moreover, the literature analysing the role of current smoking mainly neglected the potential effects of former smoking behaviours.

In the case of Indonesia, there was only one study that observed the effects of current smoking status and the number of cigarettes consumed on body weight measures (Sohn, 2015). The authors of this study did not thoroughly remove the endogeneity problems of current smoking due to leaving unobserved time-variant factors and reverse causality in the regressions. Some other drawbacks of this study were the exclusion of female samples because smoking in Indonesia was assumed to be a habit associated only with men, the unification of former smokers and non-smokers as a one group comparison with current smokers, the equalisation of chewing and piping into smoking, and the negligence of the role of smoking cessation on body weight.

I address the gap in the literature in several ways. First, I provide causal impacts of current smoking status and smoking intensity on BMI by adopting the fixed-effect model and instrumental variables. Second, as additional analysis, I examine the effects of former smoking status by partly eradicating endogeneity problems in the regressions. Third, the comparison group of current smokers is divided into former smokers and non-smokers and there is also an analysis of cigarette consumption only among current smokers. Fourth, even though women's smoking prevalence was statistically much lower than that of men in Indonesia (WHO, 2012), a sample of women is included in the study so that the effects of smoking could be analysed separately for men and women. Lastly, the thesis may contribute to the literature by providing an additional study in a developing country and filling the gap left by the earlier study in Indonesia.

3 Data and variables

3.1 Data

This research used panel data from the Indonesia Family Life Survey (IFLS), which were conducted by the RAND Corporation in collaboration with several Indonesian universities. The survey collected social and economic indicators at household and community levels, such as education, migration, health, and expenditures. There are five waves of IFLS, conducted in 1993/94 (IFLS1), 1997 (IFLS2), 2000 (IFLS3), 2007/08 (IFLS4), and 2014/15 (IFLS5) respectively. In the first wave, 7224 households and around 22,019 individuals in 13 out of 27 Indonesia's provinces were interviewed. The first wave was claimed as a sample representative of 83% of Indonesian people.

The samples used in each wave were at an individual level and restricted to respondents aged 15 years and above. The restriction was considered due to the minimum age of respondents (15 years) to answer smoking-related questionnaires in all IFLS waves. Table 1 depicts the final number of observations for those aged 15 years and above from the first until the fifth waves of IFLS after deleting duplicates and non-current household members and merging several datasets of IFLS in each wave. Overall, the total number of observations was 116,449, containing 47,217 individuals.

Table 1 The number of households and individuals interviewed across all IFLS waves

Wave	**Households**	**Individuals**	**Observations included in the study (aged 15 years and over)**
1	7,224	22,019	12,892
2	7,698	31,952	18,360
3	10,574	38,433	25,280
4	13,995	44,103	28,785
5	16,931	58,325	31,132
Total	56,422	194,832	116,449

Source: Author's compilation from IFLS data, Frankenberg & Thomas (2000), Strauss et al. (2004), Strauss, Witoelar, & Sikoki (2016), and Strauss, Witoelar, Sikoki, & Wattie (2009)

According to a study by Watanabe et al. (2016), the impact of current smoking status on body weight may have different magnitude and significance with different comparison groups, such as former smokers, non-smokers who never smoke, and non-current smokers (both former and non-smokers). Therefore, this study also separated the analysis into several samples to distinguish the reference group(s) being compared with the current smokers. Additionally, this thesis also determined the effect of smoking intensity among current smokers only. Based on some previous studies, the direction of the coefficients of smoking intensity among current smokers may be different with the ones being compared with other non-current-smoker groups (Bamia et al., 2004; Dare et al., 2015). Another sample obtained in this study contained former smokers and non-smokers to analyse the role of quitting smoking. In sum, there were five samples applied in the paper: 1) full sample, 2) current-smoker and non-smoker (CSNS) sample, 2) current-smoker and former-smoker (CSFS) sample, 3) former-smoker and non-smokers (FSNS) sample, and 4) current-smoker (CS) sample. The main smoking-related variables and the included samples can be shown in Table 2 below:

Table 2 Smoking-related variables and the included samples

Smoking-related variables	**Samples**				
	Full Sample	**CSNS**	**CSFS**	**CS**	**FSNS**
Current smoking status	✓	✓	✓		
The number of cigarettes consumed per day (smoking intensity)		✓	✓	✓	
Former smoking status	✓				✓
The number of cigarettes consumed per day before quitting smoking					✓
Duration of smoking before quitting smoking					✓
Length of time since giving up smoking					✓

I also divided the analysis for male and female because of different patterns of body weight between women and men in Indonesia (Aizawa & Helble, 2017). Also, the number of female smokers was much smaller than that of male smokers. The number of current smokers, former smokers, and non-smokers aged 15 years and above by sex in each wave are depicted in Table 3 below. Due to missing data in the sex variable (inconsistent sex data across waves was also assigned missing values), the total number of male and female observations are lower than the total individual data in Table 1. Based on the IFLS data, the percentage of female current smokers aged 15 years and above ranged from 4.72% in the first wave to 1.71% in the fifth wave, while the percentage of male current smokers ranged from 72.75% in the first wave to 63.87% in the fifth wave (see Table 3). The data was consistent with another finding that in 2013, 64.9% Indonesian males were tobacco smokers aged 15 years and above, while only 2.1% of females were tobacco smokers (Indonesian Ministry of Health, 2018).

Table 3 The number of observations for each subsample in five IFLS waves by sex (respondents aged 15 years and above)

Wave	**Current smokers**				**Former smokers**				**Non-smokers**			
	Male		**Female**		**Male**		**Female**		**Male**		**Female**	
1	4,130	72.75%	329	4.72%	529	9.32%	153	2.19%	1,018	17.93%	6,494	93.09%
2	4,966	63.44%	314	3.14%	870	11.11%	159	1.59%	1,992	25.45%	9,529	95.27%
3	7,409	64.34%	353	2.66%	1,138	9.88%	176	1.33%	2,968	25.78%	12,751	96.02%
4	8,662	65.03%	289	1.92%	1,109	8.33%	178	1.18%	3,550	26.65%	14,583	96.90%
5	9,087	63.87%	283	1.71%	1,440	10.12%	172	1.04%	3,700	26.01%	16,128	97.26%
Total	34,254	65.16%	1,568	2.53%	5,086	9.68%	838	1.35%	13,228	25.16%	59,485	96.11%

3.2 Variables

3.2.1 Body weight

This study used BMI as a measure of body weight. BMI has been widely accepted as a proxy for corpulence and slenderness, which are linked to other health problems (WHO expert consultation, 2004). This indicator is calculated by dividing body weight with the square of one's height (kg/m^2). BMI categories used in this study were based on internationally accepted BMI cut-offs for Asian people proposed by WHO expert consultation (2004). The use of Asian BMI categories rather than the international one was argued to be more appropriate for Asian countries, including Indonesia, due to higher health risks and higher fat composition of Asian people under the international normal-weight cut-off point compared with other people (WHO expert consultation, 2004). The BMI categories based on the Asian standard were underweight for BMI less than 18.5 kg/m^2, normal weight for BMI more than/equal to 18.5 kg/m^2 and less than 23 kg/m^2 (18.5 $kg/m^2 \leq$ BMI $<$ 23 kg/m^2), pre-obese for BMI more than/equal to 23 kg/m^2 and less than 27.5 kg/m^2 (23 $kg/m^2 \leq$ BMI $<$ 27.5 kg/m^2), overweight (including pre-obese and obese) for BMI more than/equal to 23 kg/m^2, and obese for BMI more than/equal to 27.5 kg/m^2.

Even though height and weight data in IFLS was obtained by trained interviewers (Strauss, Beegle, et al., 2016), there were still measurements errors. The average weight and height of Indonesian males and females at the age of 15 years (the minimum age of individuals in the study) in the literature were 47.21 kg and 156.53 cm for boys and 45.06 kg and 151.38 cm for girls (Pulungan, Madarina, Batubara, & Hermanussen, 2018). Thousands of weight and height data in the IFLS were found to be far below the average weight and height of 15-year-old boys and girls in Indonesia, with some figures being as low as 0.5 kg and 14.2 cm. I dealt with measurement errors as well as outliers, especially in the lower tails of height and weight data, by applying a two-step data management procedure: 1) assigning the values of the lowest 1% of the height and weight variables by male and female status in each wave as missing; and 2) using Grubbs method to detect outliers in Stata 15 with 99% confidence level (Grubbs, 1969). The first step assigned missing values to

around 1107 (0.97%) weight data and 1114 (0.98%) height data[3]. The second step was used to detect the remaining outliers in weight and height data for both upper and lower tails, which resulted in an additional 63 observations for the weight variable and 13 observations for the height variables being reassigned as having missing data.

3.2.2 Smoking-related behaviours

Smoking-related variables are the primary independent variables in this study and are available for current smokers and former smokers. First, as for the smoking-status variable, respondents were classified as either a non-smoker, a current smoker, or a former smoker. Respondents were classified as current smokers if they were active smokers of self-rolled cigarettes, cigarettes, or cigars at the time of the survey. Respondents were classified as former smokers if they previously had a smoking habit of self-rolled cigarettes, cigarettes, or cigars but has quit the habit by the time of the survey. Respondents were classified as non-smokers if they had never consumed self-rolled cigarettes, cigarettes, or cigars. Smoking status does not account for piping, shisha, vape, electronic cigarettes, and chewing tobacco. First, shisha, vape, and electronic cigarettes are not accounted for in IFLS data. Second, chewing tobacco was not categorised as smoking but as smokeless tobacco (O'Connor, 2012). Lastly, even though, internationally and nationally in Indonesia, piping tobacco was categorised as smoking (O'Connor, 2012; Statistics Indonesia, 2017), it was not categorised as smoking in my analysis to provide a consistent estimate in other smoking-related variables. Currently, in IFLS, there is no count of the number of pipes smoked as there is for the number of sticks of cigarettes smoked (another smoking-related variable). Moreover, piping only accounts for 0.2% of observations across all waves. Therefore, excluding this indicator was expected to have a negligible effect on the analysis. For the rest of this study, *smoking* means smoking self-rolled cigarettes, cigarettes, or cigars. That is, respondents are still classified as non-smokers if they are in the habit of smoking/using a pipe, electronic cigarettes, shisha, vape, and smokeless tobacco (chewing).

[3] The missing values were assigned to 1% of the data of 10 subsamples (five waves divided by sex), so that the total assigned values for all waves and samples might not be exactly 1% (also considering several missing values for sex).

Another smoking-related variable used in this study is the number of cigars/cigarette sticks consumed per day. According to previous studies, the magnitude and significance of the impacts of smoking status and daily cigarette consumption on BMI may be different (Sohn, 2015; Wang, 2015; Watanabe et al., 2016). The analysis on cigarette consumption allowed researchers to group smokers according to their levels of consumption—occasional, moderate, and heavy smokers (Z. Chen et al., 2007). I also included a variable noting the number of cigarettes smoked daily by current smokers to measure the effects of smoking intensity on BMI. This variable was a self-reported average consumption of cigars/cigarettes per day at the time of interview. Dissimilar to a study by Sohn (2015), I did not transform the tobacco (in grams) consumed through chewing and piping into the number of cigarette sticks. Furthermore, the measure used by Sohn to convert 1 gram of tobacco consumed per day into one stick of cigarettes may be ill-founded. Therefore, I only measured smoking intensity from cigars/cigarette and excluded tobacco consumption from piping and chewing to ensure the pure effects of cigarette smoking intensity. I also made a duration of smoking variable for current smokers by subtracting the age when they started smoking from their age at the time of interview[4].

There were three additional indicators related to former-smoking behaviours: 1) smoking intensity before quitting smoking, 2) duration of smoking behaviour, and 3) length of time since quitting smoking. The duration of smoking for former smokers was constructed by subtracting the age when they started smoking from the age when they totally quit smoking. There is an additional variable for former smokers only, namely the length of time since quitting smoking, to capture the impact of smoking absence after being smokers on body weight. The variable was calculated by subtracting the age when former smokers quit smoking from their age at the time of interview. The number of cigars/cigarettes consumed per day, duration of smoking, and duration of quitting smoking for former smokers has a value of 0 (in cigarette and year respectively) for non-smokers.

[4] The variable ended up being removed from the analysis due to the absence of valid instruments which will be explained in section 4 (Methods).

3.2.3 Other covariates

Both individual- and household-level control variables were used in this study. At the individual level, these were age, age square, marital status, years of schooling, informal job status, and total hours of work per week. Age was expected to have positive impacts on body weight (Baum, 2009; Fang et al., 2009; Plurphanswat & Rodu, 2014; Sohn, 2015; Wehby et al., 2012) as older people tend to be less active and have positive energy balance (Jura & Kozak, 2016), while age square was found to be negatively correlated with body weight (Z. Chen et al., 2007; Plurphanswat & Rodu, 2014; Sohn, 2015). Marital status had a value of 1 for married and 0 for not married, separated, widow, and cohabitating. The status of cohabitation is not a legal status in Indonesia and accounts for only 0.01% of total observations, so people who were cohabitating were grouped with those who were single. Most previous studies found that married people tend to have higher body weight (Baum, 2009; Fang et al., 2009; Guerra et al., 2015; Plurphanswat & Rodu, 2014; Sohn, 2015; Wehby et al., 2012). According to the proposed relationship between education and health by Grossman (1972), higher level of education was expected to promote better health, including a healthy body weight (Cawley, 2015). However, previous studies found mixed results regarding the correlation between education and body weight or BMI (Z. Chen et al., 2007; Fang et al., 2009; Plurphanswat & Rodu, 2014; Wehby et al., 2012). I used the total years of schooling as a proxy of education, which was found to have a positive impact on body weight or BMI (Baum, 2009). The total hours of work per week were included to control the impact of working on body weight. The effect of total working hours on body weight may vary depending on whether a job was strenuous or non-strenuous (Abramowitz, 2016). Abramowitz (2016) found that working hours positively affected the BMI of those in non-strenuous jobs but did not affect BMI of those in strenuous jobs. However, I did not include the interaction between job status and total working hours because I was analysing the impact of smoking on both working and non-working populations.

My research also employed a dummy variable of urban location, a dummy variable of provinces in Java Island, and the natural log of effective per capita income as control variables at the household level. Previous studies showed that residing in urban areas is positively related to body weight (Baum, 2009; Fang et al., 2009; Sohn, 2015). The result may be caused by the wider availability of fast food outlets

and urban residents' liking of eating out (Cawley, 2015). Several studies used provinces as one of the predictors of body weight (Z. Chen et al., 2007; Fang et al., 2009). The dummy variable of provinces in Java Island was used as the island takes up more than half of Indonesia's economy and is claimed to be the most developed region of Indonesia (Indonesia Investments, 2013). The dummy variable had a value of 1 if the household resided in Banten, DKI Jakarta, DI Yogyakarta, West Java, Central Java, or East Java and a value of 0 if the household resided elsewhere. Another household-level variable included in the regression was the log of effective per capita income. Household income was obtained by summing up household members' incomes, and effective per capita income was derived by dividing household income with the square root of household members to account for within household resources sharing. Theoretically, the relationship between income and BMI or body weight can either be positive or negative: it will be positive if one values foods and sedentary leisure higher than health and appearance (Cawley, 2015). However, previous studies found no conclusive evidence regarding the impact of income per capita or individual monthly earnings on body weight (Z. Chen et al., 2007; Fang et al., 2009; Sohn, 2015).

It has been found that the effect of smoking on body weight may be mediated by consumption (Wehby & Courtemanche, 2012), since cigarettes smoking may decrease available resources to consume other foods and non-food consumption (R. M. John, 2008; Jumrani & Birthal, 2017; Pu, Lan, Chou, & Lan, 2008). Therefore, I also analysed the effect of additional household per capita sugar and oil consumption variables on body weight to see whether the coefficients of smoking-related variables may be affected by the inclusion of these consumption variables. Consumption of sugar and oil in IFLS was determined by the quantity of granulated sugar (in kilograms or litres) and cooking oil (in litres) purchased last time, if the household bought any granulated sugar or cooking oil within the last month. Household per capita sugar and oil consumption was derived from household granulated sugar and cooking oil consumption divided by the square root of household members. Previous studies found that household/individual food consumption tended to increase body weight (Z. Chen et al., 2007; Fang et al., 2009; Wehby & Courtemanche, 2012), but the impact of the addition of the regressors on smoking-related variables was still unobserved.

3.3 Descriptive statistics

3.3.1 Descriptive statistics by sex

Table 4 presents the descriptive statistics stratified by sex and definitions of the variables employed in this study. Mean weight and height of males was significantly higher than females. However, the difference in average height between males and females was more than twice higher than the difference in their average weight. As a result, even though both males and females had a normal BMI, on average, the mean BMI of males was slightly lower than that of their female counterparts (21.46 kg/m^2 vs 22.84 kg/m^2). The finding is consistent with Ritchie & Roser (2019), who showed that the average BMI of Indonesian women was 1.04 points higher than that of men in 2014 (23.36 kg/m^2 compared to 22.32 kg/m^2). There were minimal variations of BMI among current smokers, former smokers, and non-smokers for both sexes (around 21 kg/m^2 for men and 22 kg/m^2 for women). Compared with women, men were significantly more likely to have a healthy weight (52.34% vs 42.54%) and less likely to be in the pre-obesity category (20.08% vs 27.12%) as well as less likely to be in obesity category (10.18% vs 17.12%). Overall, based on the IFLS data, there seemed to be different means for weight, height, BMI, and BMI status between men and women in Indonesia.

The percentage of males who were current or former smokers was around 25 and 7 times higher than that of female (65.16% vs 2.53% and 9.68% vs 1.35%). The finding was in line with those from other sources, for example, the World Bank Group (2018) reported that Indonesia's percentage of male current smokers was around 27 times or 65.6 percentage points higher than the percentage of female current smokers in 2015. As current smokers, men on average had been smoking in shorter periods than women. Male current smokers smoked between 3 to 4 more cigarettes per day than their female counterparts, while former-smoking men smoked approximately 5 to 6 more cigarettes per day than former-smoking women before they quit smoking. Amongst former smokers, the duration of their previous smoking behaviour was around 1.5 times longer for males than for females, but the length of time since quitting smoking (i.e. the year since they gave up smoking till the year of the interview) was not statistically different between males and females.

The average ages of men and women in our sample were very much the same at around 37 years. Although, men were slightly more likely than women to be married (71.38% vs 69.07%). The total working hours per week of men was almost twice that of women due to a much higher proportion of men entering the labour force than women in Indonesia (approximately 80% of men and 50% of women in 2016) (ILO, 2017). The total years of schooling was around one year higher for males. Men and women tended to have typical household- and community-level characteristics with only small (yet statistically significant) differences regarding the instrumental variables, per capita income, and per capita consumption.

3.3.2 Descriptive statistics by smoking status

This section provides the descriptive statistics for the variables that indicate the three possible smoking statuses of respondents: current smoker (CS), former smoker (FS), and non-smoker (NS). Importantly, Table 5 shows that almost all current smokers were males (95.62%), while around a quarter of former smokers and more than 81% of non-smokers were females. These figures are essential for understanding the distribution of BMI.

As seen in Table 5, the mean weight and height of non-smokers were significantly lower than the other two groups, while their mean BMI was higher. This finding is probably because a large proportion of non-smokers is female. However, the mean BMI of former smokers was slightly higher than current smokers. The distribution of non-smokers' BMI seemed to be similar to the distribution of females' BMI (as shown in Table 4) due to, again, a higher portion of non-smokers being females. For the same reason, the distribution of BMI among current smokers followed that of males.

Among the three groups, the group of former smokers had the highest mean age, indicating that former smokers tended to quit at a late age. Among current smokers, the higher proportion of those married and the longer working hours per week may be due to the higher proportion of men in this group, as observed in Table 4. Even though non-smokers were mostly women, who had lower mean years of schooling as seen in Table 4, the years of schooling of non-smokers was slightly higher than current smokers. This denotes that unhealthy behaviours, such as smoking, may be associated with lower education (Amalia, Cadogan, Prabandari, &

Filippidis, 2019; Grossman, 1972). In line with a study by Amalia et al. (2019) in Indonesia, the mean years of education of former smokers was higher than that for current smokers. The three groups shared similar household characteristics regarding location, income, and consumption, with only slight significant differences.

Table 4 Descriptive statistics of analysis variables by sex (respondents aged 15 years and above)

Variable names	Variable descriptions	Male		Female		Mean Diff.
		Mean	SD	Mean	SD	
Body measurements						
Weight (kg)	Body weight in kg	56.35	10.46	51.73	10.57	4.617***
Height (cm)	Body height in cm	161.84	6.20	150.46	5.54	11.38***
BMI	BMI (weight/height-squared)	21.46	3.46	22.84	4.28	-1.371***
BMI of current smokers	BMI of current smokers	21.24	3.26	22.27	4.44	-1.028***
BMI of former smokers	BMI of former smokers	21.92	3.81	22.52	4.62	-0.596***
BMI of non-smokers	BMI of non-smokers	21.86	3.76	22.85	4.27	-0.990***
Underweight	(1) if BMI less than 18.5 kg/m^2; (0) otherwise	17.39%	0.38	13.22%	0.34	0.0418***
Normal weight	(1) if 18.5 $kg/m^2 \leq$ BMI $<$ 25 kg/m^2; (0) otherwise	52.34%	0.50	42.54%	0.49	0.0980***
Pre-obese	(1) if 23 $kg/m^2 \leq$ BMI $<$ 27.5 kg/m^2; 0 otherwise	20.08%	0.40	27.12%	0.44	-0.0704***
Obese	(1) if BMI was equal to or more than 27.5 kg/m^2; (0) otherwise	10.18%	0.30	17.12%	0.38	-0.0693***
Current smoking						
Current smokers	(1) if currently smoke self-rolled cigarettes, cigarettes, or cigars; (0) otherwise	65.16%	0.48	2.53%	0.16	0.626***
Daily cigarettes smoked by current smokers	The number of average daily cigarettes smoked by current smokers	11.26	6.72	7.34	6.07	3.913***
Smoking duration of current smokers	Duration of smoking of current smokers from the age of starting smoking till the year of interview	20.62	14.38	23.01	16.44	-2.388***
Former smoking						
Former smokers (FS)	(1) if once was an active smoker of self-rolled cigarettes, cigarettes, or cigars, but has given up the habit; (0) otherwise	9.68%	0.30	1.35%	0.12	0.0832***
Daily cigarettes smoked by former smokers	The average number of daily cigarettes former smokers used to smoke	12.35	9.16	6.51	6.47	5.838***
Smoking duration of former smokers	Duration of former smoking (in years)	21.95	15.62	16.46	14.59	5.487***

Duration of smoking absence of former smokers	Duration since quitting smoking till the year of the interview (in years)	8.91	9.24	8.73	9.70	0.178
Instrumental variables						
Percentage of current smokers in the community	Percentage of current smokers in a community (*kabupaten/kota*)	31.94%	3.23	31.83%	3.27	0.106***
Community's average daily cigarettes smoked of current smokers	Average number of daily cigarettes smoked by current smokers in a community (excluding the respondent's cigarette smoked)	11.16	1.01	11.18	1.03	-0.0150*
Other individual variables						
Age	Age of the respondent	37.87	16.06	37.67	16.00	0.203*
Married	(1) if married; (0) otherwise	71.38%	0.45	69.07%	0.46	0.0231***
Hours of work	Total working hours per week (0 for unemployed and those not in the labour force)	41.47	28.73	22.91	28.01	18.57***
Years of schooling	Total years of schooling since starting primary school	8.23	4.24	7.30	4.57	0.928***
Other household variables						
Urban	(1) if living in an urban area; (0) otherwise	51.79%	0.50	52.20%	0.50	-0.00409
Java	(1) if living in a province on Java Island (Banten, DKI Jakarta, DI Yogyakarta, West Java, Central Java, or East Java); (0) otherwise	57.77%	0.49	57.28%	0.49	0.00484
Log effective per capita income	Log of effective per capita income (household income divided by the square root of household members)	12.72	1.49	12.59	1.54	0.122***
Log effective per capita oil cons.	Log of effective per capita cooking oil consumption (the amount of cooking oil bought last time within a month divided by the square root of household members)	9.08	0.95	9.10	0.95	-0.0234***
Log effective per capita sugar cons.	Log of effective per capita granulated sugar consumption (the amount of granulated sugar bought last time within a month divided by the square root of household members)	9.07	0.93	9.08	0.94	-0.0155**
Observations		52568		61891		114459

*$P < 0.1$, **$P < 0.05$, ***$P < 0.01$

Table 5 Descriptive statistics of analysis variables by sex (respondents aged 15 years and above)

Variables	Current smokers (CS)		Former smokers (FS)		Non-smokers (NS)		Mean diff. CS vs FS	Mean diff. CS vs NS	Mean diff. FS vs NS
	Mean	SD	Mean	SD	Mean	SD			
Body measurements									
Weight (kg)	55.50	10.05	56.20	11.72	52.90	10.92	-0.693***	2.610***	3.303***
Height (cm)	161.27	6.58	159.47	7.37	152.67	7.30	1.800***	8.600***	6.799***
BMI	21.30	3.33	22.02	3.95	22.68	4.20	-0.716***	-1.374***	-0.658***
Underweight	17.06%	0.38	17.34%	0.38	13.77%	0.34	-0.00282	0.0329***	0.0357***
Normal weight	54.63%	0.50	43.75%	0.50	43.28%	0.50	0.109***	0.114***	0.00473
Pre-obese	18.02%	0.38	23.83%	0.43	26.66%	0.44	-0.0581***	-0.0864***	-0.0283***
Obese	10.29%	0.30	15.08%	0.36	16.30%	0.37	-0.0479***	-0.0600***	-0.0121*
Instrumental variables									
Percentage of current smokers	32.21%	3.28	31.69%	3.32	31.73%	3.23	0.521***	0.486***	-0.0353
Community's daily cigarettes	11.14	1.02	11.23	1.00	11.19	1.02	-0.0957***	-0.0488***	0.0469***
Other individual variables									
Male	95.62%	0.20	85.85%	0.35	18.19%	0.39	0.0977***	0.774***	0.677***
Age	39.38	15.28	43.38	18.93	36.54	15.93	-4.005***	2.831***	6.836***
Married	76.19%	0.43	71.25%	0.45	67.10%	0.47	0.0493***	0.0909***	0.0415***
Hours of work	44.37	28.13	32.18	29.12	25.10	28.59	12.19***	19.27***	7.086***
Years of schooling	7.60	4.26	7.85	4.52	7.77	4.52	-0.248***	-0.171***	0.0775
Other household variables									
Urban	48.16%	0.50	56.55%	0.50	53.69%	0.50	-0.0839***	-0.0554***	0.0285***
Java	57.17%	0.49	57.28%	0.49	57.49%	0.49	-0.00107	-0.00313	-0.00206
Log effective per capita income	12.60	1.50	12.67	1.54	12.67	1.52	-0.0629**	-0.0643***	-0.00134
Log per capita oil cons.	9.03	0.94	9.14	0.95	9.12	0.95	-0.105***	-0.0872***	0.018
Log per capita sugar cons.	9.03	0.93	9.13	0.96	9.09	0.94	-0.0964***	-0.0597***	0.0367**
Observations	37150		6119		73180		43269	110330	79299

*$P < 0.1$, **$P < 0.05$, ***$P < 0.01$

4 Methods

In order to make a causal inference from a regression estimation, one needs to account for any endogeneity problems due to omitted variable bias, reverse causality, and measurement errors. Endogeneity problems can be dealt with the application of one of two methods: 1) implementing the fixed-effect or first-difference model, or 2) using a two-stage least square estimation (Wooldridge, 2016). I combined both the two-stage least square and fixed-effect methods to deal with endogeneity problems and to thus draw causal inference from smoking to body weight. The combination of these two approaches has been used in health economics research to address endogeneity from unobserved heterogeneity (by using fixed-effect models) and other sources of endogeneity, such as time-variant errors, reverse causality, and measurement errors (by implementing two-stage least square) (Liang & Mirelman, 2014; Milner et al., 2017; Pan, Lei, & Liu, 2016; Tekin & Markowitz, 2008).

4.1 Endogeneity of smoking and fixed-effect model

The relationship between smoking and body weight can be represented by the following model:

$$BMI_{it} = \beta_1 S_{it} + \beta_2 X_{it} + \alpha_i + u_{it}, \quad (1)$$

where BMI_{it} is the body mass index of individual i at time t, S_{it} is current smoking and former smoking behaviours, α_i is unobserved heterogeneity or time-invariant factors, and u_{it} is time-variant factors. In the relationship between smoking and body weight, unobserved heterogeneity α_i is likely correlated with smoking behaviours, S_{it}, which created endogeneity problems due to omitted variable bias. For example, unobserved childhood characteristics, which tend not to vary over time, like physical, verbal, and sexual abuse in a household and living with alcoholic, mentally-ill, or drug-addicted household members may increase the likelihood of both smoking and obesity (Davis et al., 2019; Ford et al., 2011; Rehkopf et al., 2016). Additionally, a respondent may have an unobserved *addictive* personality characteristic that promotes both overeating and smoking behaviours, or an *impatient* temperament whereby they do not value future health, resulting in current unhealthy behaviours (Cawley, Markowitz, & Tauras, 2006). The presence of unobserved heterogeneity was also observed in former smoking by other researchers

(Au et al., 2013). If endogeneity from unobserved heterogeneity does occur, the random effect model is biased because it assumes strict exogeneity of no correlation between S_{it} and α_i (Wooldridge, 2016). On the other hand, the fixed-effect model is more robust than its random effects counterpart because it removes unobserved heterogeneity α_i through time demeaning.

I conducted a Hausman fixed-vs-random effects test to see whether the strict exogeneity assumption was violated (Hausman, 1978). The test was implemented after panel regressions of the full sample with the random effect option by using the user-free contribution command called `xtoverid` in Stata 15 to allow for heteroskedasticity-robust standard errors (Schaffer & Stillman, 2010). The test results presented in Appendix A.1-A.4 show a p-value = 0.0000, suggesting to reject the null hypothesis of no correlation between the explanatory variables and unobserved heterogeneity for all estimations with different endogenous variables. Based on the previous literature and the results of the Hausman test, I applied FE in all regressions to address the endogeneity problems caused by unobserved heterogeneity.

4.2 Current-smoking behaviours: two-stage least square with fixed effects

The fixed-effects model may be able to remove endogeneity due to unobserved heterogeneity or the time-invariant errors. However, current smoking may still be endogenous due to the correlation between current-smoking-related variables (CS_{it}) and the time-varying errors (u_{it}), as well as reverse causality between current-smoking behaviours and body weight. For example, lower self-rated health and the presence of post-traumatic stress disorder, which can vary over time, has been found to increase the likelihood of both smoking and obesity (Berk-Clark et al., 2018; Mendoza-Romero et al., 2019). Furthermore, reverse causality may arise because smoking was found to be one of the weight-control methods and a person with a higher weight is more likely to be a smoker (Cawley et al., 2006). I applied the two-stage least square methods to address these remaining potential endogeneity problems which cannot be solved by the fixed-effects method.

An endogeneity test was conducted to confirm the endogeneity of current-smoking behaviours. According to Wooldridge (2016), the endogeneity test can be performed by including residuals from the first stage as an additional regressor in the structural equation (1). If the coefficient of the residual from the first stage regression is statistically different from zero in the t-test, then the instrumented variable is likely to be endogenous (i.e. rejecting the null hypothesis that the instrumented variable is exogenous). The test was run using Stata `xtreg` command for all current-smoking behaviours, including current-smoking status because Stata `xtprobit` command cannot predict the residuals. To show the necessity of the two-stage least square (to assess if there is endogeneity problem due to time-variant errors), the residuals obtained from the first stage regression captured only time-variant errors (using the `e` option after `predict` command) and was included as an additional explanatory variable to the random-effect structural equation with observations from the full sample. All coefficients of residuals in three different structural equations were statistically different from zero, so that current smoking behaviours, average daily cigarettes smoked by current smokers, and duration of smoking of current smokers were all endogenous (see Appendix B for various endogeneity test results).

4.2.1 Instrumental variables (IVs)

The instrumental variables (IVs) I used in this study were measures of community-average smoking behaviours. These measures included (1) the percentage of people in the community being current smokers and (2) the average number of cigarettes smoked by current smokers per day in the community. I also attempted to apply two other instruments for duration of smoking of current smokers: (a) the average number of years smoking for the current smokers in the community and (b) for current smokers, the total number of years being exposed to the smoking culture in their community, which is calculated as the person's age multiplying with measure (b).

The level of community that I studied was *kabupaten/kota*, which is equivalent to a district or city. I chose this level of community because many lower-level communities in IFLS—*kecamatan* (sub-district) or *kelurahan* (village-level

community)—contained very few observations. Following Fang, Ali, & Rizzo (2009), all the instrumental variables excluded the respondent's value of the related indicators to ensure a true measure of the community's smoking behaviours. Similar studies in China had also used the number of cigarettes smoked at the community level as an IV for the number of cigarettes smoked (Fang et al., 2009) and had used province-level of cigarette production as an IV for current smoking status and the number of cigarettes consumed daily (Wang, 2015). Moreover, area-based instruments have been widely used by other studies in health economics research (Grabowski & Hirth, 2003; Morris, 2006, 2007; Sasso & Buchmueller, 2004; Sloan, Picone, Taylor Jr., & Chou, 2001).

A variable must fulfil two requirements to be an IV: 1) *relevance* or correlated with the endogenous regressor ($cov(Z_{it}, X_{it}) \neq 0$); and 2) *exogeneity* or uncorrelated with the error term in the structural equation ($cov(Z_{it}, u_{it}) \neq 0$) (Wooldridge, 2016). Regarding the *relevance* criterion, the instruments that were weakly correlated to the endogenous variables may cause inconsistencies as well as the huge and unexpected sign of coefficients of the endogenous variables in the second stage, even when the *exogeneity* criterion is met (Bound, Jaeger, & Baker, 1995; Wooldridge, 2016). Therefore, I followed the recommendation by Stock & Yogo (2005) on how to detect weak instruments. An instrument can be accepted if the instrument is statistically significant at 5% significance level when there is only one instrument and, in the case of multiple instructions, when the test statistic from a joint-significance F test at the first-stage regression is larger than 10[5].

The results in Appendix C depicted the fulfilment of the *relevance* requirement of the instruments used in this study. The absolute values of the t-statistics for the percentage of current smokers in the community and the community's average number of cigarettes smoked daily by current smokers was larger than 3.2, indicating that they were acceptable instruments for current smoking status and daily cigarettes smoked by current smokers. Meanwhile, the F-statistics of the community's average duration of current smoking of current smokers and the respondent's age exposure to the smoking culture in the community was far less than 10, meaning that they could not be used as instruments for the duration of smoking

[5] this is not the F-statistics from all exogenous variables, but only from the coefficients of instruments (Wooldridge, 2016).

of current smokers. As a result, the duration of smoking was *excluded* from the analysis because of the lack of valid instruments.

The second requirement of instruments, *exogeneity* ($cov(Z_{it}, u_{it}) \neq 0$), cannot be tested unless there is more than one instruments (Wooldridge, 2016). An overidentification test can be conducted if there is more than one instruments employed. In this case, the test was not obtainable for the percentage of current smokers in the community and the community's average number of cigarettes smoked daily by current smokers since each of the instruments was used for one endogenous variable (the percentage of current smokers in the community as an IV for the current smoking status; and the community's average number of cigarettes smoked daily by current smokers as an IV for the number of cigarettes smoked daily by current smokers).

There may be arguments for using other variables as instruments for current-smoking behaviours. There were variables in IFLS that could potentially have been used as instruments. These were: 1) cigarette prices; and 2) parent's smoking status when the respondent was at the age of 12. The cigarette price of the most-sold cigarette brand was used in Fang et al. (2009) as an IV of the number of cigarettes smoked daily by current smokers. It is argued that cigarette price may be correlated with body weight through channels other than smoking, such as the consumption of other foods, so that cigarette price may also be endogenous (Wehby & Courtemanche, 2012). Another possible IV, namely parents' status of smoking when the respondent was at the age of 12, may have a problem of recall bias because people are asked about their parents' smoking status when they were a particular age (12 years old) instead of during a broader age range (e.g. between the ages of 5 and 18 years) (Webb & Bain, 2011). Moreover, since this variable was only available in the fifth wave, only individuals who appeared in the fifth wave could have been included in the study if this variable were to be used as an IV. Thus, even though there was evidence on the impact of parents' smoking behaviours on offspring (Melchior, Chastang, Mackinnon, Galéra, & Fombonne, 2010), I opted not to use this variable as an instrument. To conclude, the community-level instruments were the best IVs among all the options available from IFLS data.

4.2.2 First-stage regressions

A first stage regression was run separately for three different current-smoking-related variables. The number of daily cigarettes smoked by current smokers was estimated by panel data model with fixed-effects (using `xtreg` command) because the Hausman test rejected the null hypothesis of no correlation between explanatory variables and time-invariant errors (see Appendix A.5.). Meanwhile, provided that there was no way to run a Hausman test to choose random effects of FE for panel probit or logit model, this study decided to estimate current smoking status by the probit model with random effects (using the `xtprobit` command).

There were some considerations when using a random effects probit model to estimate current smoking status. First, the FE logit model is only available for switchers and excludes non-switchers (i.e. respondents being current smokers or non-smokers in all periods) (Cameron & Trivedi, 2005), so that a very significant number of observations (e.g. about more than seventy nine thousand observations out of around eighty nine thousand observations of both sexes in the full sample) were dropped. To provide a similar estimation of population parameters as in other first-stage and second-stage regressions, this model was not employed to estimate the first-stage regression of current smoking status.Second, there were also efforts to replace the fixed-effects model by applying a random effects with Mundlak corrections, since the method was claimed to produce similar results with the ones with FE (Mundlak, 1978). However, both panel probit and logit failed to calculate robust-standard errors and p-values for some variables in the male CSNS[6] sample (the standard errors and p-values appeared as missing), so that the coefficient values might be different with the regression results which included a full set of variables. Therefore, probit random effect models were obtained to estimate the first-stage regressions of current smoking status for all samples. The first-stage regressions for two current-smoking-related variables can be shown in the following equations:

$$\Pr(CS\_status_{it} = 1 \mid Z_{1it},\ X_{it}) = \alpha_1 Z_{1it} + \alpha_2 X_{it} + \theta_i + \varepsilon_{it} \quad (2)$$

$$CS\_cig_{it} = \alpha_1 Z_{2it} + \alpha_2 X_{it} + \theta_i + \varepsilon_{it} \quad (3)$$

[6] CSNS: current smokers and non-smokers.

where CS_status_{it} was current smoking status of individual i at time t, CS_cig_{it} was the current average number of cigarettes smoked daily, Z_{1it} was an instrumental variable for current smoking status (the percentage of currents smokers in the community), Z_{2it} was an instrumental variable for the number of cigarettes smoked daily by current smokers (community average of daily cigarettes smoked by current smokers), X_{it} was a set of control variables, θ_i was unobserved heterogeneity, and ε_{it} was time-variant error.

4.2.3 Structural equation: the relationship of current smoking and body weight

The second-stage regressions in this study were estimated by using fixed-effects models. The estimations were run for four different samples: 1) the full sample with current smokers, former smokers, and non-smokers; 2) a sub-sample with current smokers and non-smokers (CSNS); 3) a sub-sample with current smokers and former smokers (CSFS); and 4) a sub-sample with current smokers only (CS). A separate regression by gender was to test if there were different impacts of smoking on body weight in men and women. The structural equations are:

$$BMI_{it} = \beta_1 CS\_status_{it} + \beta_2 FS\_status_{it} + \beta_3 X_{it} + \alpha_i + u_{it} \quad (4)$$

$$BMI_{it} = \beta_1 CS\_status_{it} + \beta_3 X_{it} + \alpha_i + u_{it} \quad (5)$$

$$BMI_{it} = \beta_1 CS\_cig_{it} + \beta_3 X_{it} + \alpha_i + u_{it} \quad (6)$$

where FS_status_{it} was former smoking status of individual i at time t, α_i was unobserved heterogeneity, and u_{it} was time-variant error. Equation (4) was only applied in the full sample to test the impact of both current and former smoking status on body weight. On the other hand, equation (5) was estimated in both CSNS and CSFS sub-samples, while equation (6) was run in CSNS, CSFS, and current-smoker-only (CS) sub-samples.

4.3 Estimations of former-smoking behaviours

This study also investigated the relationship between former-smoking behaviours and body weight by partly removing the problem of endogeneity with

fixed-effects models. As suggested in the literature, time-variant factors and reverse causality may also be the sources of endogeneity of former smoking (Amir, 1996; Bush et al., 2014; Reid & Ledgerwood, 2015). Additionally, some previous economics papers also treated former smoking as endogenous and removed the problem of endogeneity by applying instrumental variables and difference-in-difference methods (Courtemanche et al., 2018; Kasteridis & Yen, 2012; Pieroni & Salmasi, 2016). Unfortunately, there seemed no instruments could be used in the study to isolate the effects of former smoking on body weight by using IFLS data. Medical intervention from a randomised control trial and cigarette tax used by previous studies as instruments of former smoking were undoubtedly unable to be obtained from the data. Moreover, it appeared that there was no community-level instrumental variable which can be observed for former smoking. For instance, there was no possible way to create the community's percentage of former smokers at the time the respondents gave up smoking since quitting smoking could take place outside IFLS interview periods. Consequently, similar to a study by Pieroni & Salmasi (2016), I only removed partial endogeneity problems originating from the unobserved heterogeneity. A fixed-effects method *without* instrumental variables was applied to examine the role of former-smoking behaviours. The use of a fixed effects method was supported by the Hausman test showing the rejection of the null hypothesis of no correlation between explanatory variables and the unobserved heterogeneity (see Appendix A.6). For this reason, the analysis of the effects of former smoking on body weight may have limitations on causal inference.

There were four former-smoking related variables considered in this study: 1) former smoking status, 2) the number of cigarettes smoked daily before quitting smoking, 3) the duration of smoking before quitting smoking, and 4) length of time since giving up smoking. The following equations display the correlation between the former-smoking variables and BMI:

$$BMI_{it} = \rho_1 FS\_status_{it} + \rho_2 X_{it} + \alpha_i + u_{it} \quad (7)$$

$$BMI_{it} = \rho_1 FS\_cig_{it} + \rho_2 X_{it} + \alpha_i + u_{it} \quad (8)$$

$$BMI_{it} = \rho_1 FS\_duration_{it} + \rho_2 X_{it} + \alpha_i + u_{it} \quad (9)$$

$$BMI_{it} = \rho_1 FS\_quit_{it} + \rho_2 X_{it} + \alpha_i + u_{it} \quad (10)$$

where FS_cig_{it} was former daily cigarettes smoked, $FS\_duration_{it}$ was former duration of smoking, FS_quit_{it} duration of quitting smoking, X_{it} was a set of control variables, α_i was unobserved heterogeneity, and u_{it} was time-variant error. Equation (7) to (10) were all estimated in former-smoker and non-smoker sub-sample (FSNS).

5 Results

5.1 First-stage regression

5.1.1 First-stage regression: current smoking status

The first-stage regression for current smoking status was estimated for full, CSNS (current plus non-smokers), and CSFS (current plus former smokers) samples. As depicted in Table 6, the percentage of current smokers in the community (the instrumental variable) had a positive impact on the probability of being a current smoker in all the three samples. Even though the IV may only increase the probability of becoming a current smoker by about 0.1 to 1 percentage point across all samples, all coefficients were statistically significant at the 1% level of significance. It can be inferred from Table 6 that when the percentage of current smokers in the community increases by one percentage point, the likelihood of being a current smoker increases by around 0.5-0.6 percentage points, on average, irrespective of gender, by around 0.4-0.8 percentage points, on average, for men, and by around 0.1-1 percentage points, on average, for women. The full estimation results of the first-stage regression of current smoking status can be seen in Appendix D.1-D.3.

5.1.2 First-stage regression: average number of cigarettes smoked daily by current smokers

The first-stage regression of the number of daily cigarettes smoked by current smokers was run for the CSNS, CSFS, and CS samples. As can be seen in Table 6, community's average daily cigarette consumption did not significantly affect the number of cigarettes smoked daily by female current smokers (except for the CS sample) but had significant impacts for both sexes together and for male groups in all the three samples. Even though the IV affected the number of cigarettes smoked daily by females in the CS sample, the coefficient was only significant at the 10%

significance level, which is below the threshold suggested by Stock & Yogo (2005) for a strong IV[7]. Based on this result, the second stage regression of the related endogenous variable was not conducted for the female cohort for any samples. The full estimation results of the first-stage regressions of the number of cigarettes smoked daily by current smokers can be seen in Appendix D.4-D.6.

Overall, for every cigarette stick increase in the community's average daily cigarette consumption, the cigarette consumption of both sexes and male current smokers may increase by around 0.11 to 0.35 stick, on average. A similar pattern to the first-stage regression results was also observed in a study by Fang, Ali, & Rizzo (2009), in which male's consumption of cigarettes potentially increased by 0.403 stick per day, on average, for every one cigarette increase in the average number of cigarettes smoked in the local community. The first-stage regression results of this study also inferred that, in the CSNS sample, every 10-stick increase in the community's average daily consumption of cigarettes smoked was associated with an increase of an individual's daily cigarette consumption by around 1.1, on average, for both sexes and by around 2.3, on average, for males. Whereas in the CSFS sample, every 10-stick increase in the community's average daily consumption of cigarettes smoked was associated with an increase of individual daily cigarette consumption by around 1.5, on average, for both sexes and by around 1.6, on average, for males. The impact tended to be larger for the CS sample where a 10-stick increase in the community's average daily cigarette consumption may potentially add around 3.4 cigarette sticks per day, on average, for both sexes and male current smokers.

[7] Stock & Yogo (2005) recommended that at least an IV is significant at 5% significance level to be a strong IV.

Table 6 Selected coefficients of the first-stage regression results

Instrumental variables	**Endogenous variable: current smoking status Average marginal effects of first-stage regressions with probit and random effects**								
	Full sample			**Current-smoker and non-smoker (CSNS) samples**			**Current-smoker and former-smoker (CSFS) samples**		
	Both sexes	**Male**	**Female**	**Both sexes**	**Male**	**Female**	**Both sexes**	**Male**	**Female**
Percentage of current smokers in the community	0.00551***	0.00790***	0.00108***	0.00553***	0.00377***	0.00105***	0.00502***	0.00470***	0.0140***
	(0.000487)	(0.000761)	(0.000245)	(0.000643)	(0.000728)	(0.000247)	(0.000698)	(0.000699)	(0.00464)
Observations	89507	41730	46632	84904	37761	46099	33497	31092	1542

Instrumental Variables	**Endogenous variable: the number of cigarettes smoked daily by current smokers First-stage regression with fixed effects**								
	Current-smoker and non-smoker (CSNS) samples			**Current-smoker and former-smoker (CSFS) samples**			**Current-smoker (CS) sample**		
	Both sexes	**Male**	**Female**	**Both sexes**	**Male**	**Female**	**Both sexes**	**Male**	**Female**
Community's Average Daily Cigarettes smoked by current smokers	0.106***	0.230***	0.00949	0.153**	0.162**	-0.0191	0.339***	0.346***	0.513*
	(0.0228)	(0.0521)	(0.00644)	(0.0627)	(0.0654)	(0.233)	(0.0654)	(0.0673)	(0.309)
Observations	84487	37377	46079	33058	30688	1520	28480	26741	990

Cluster-robust standard errors in parentheses

$^{*}P < 0.1$, $^{**}P < 0.05$, $^{***}P < 0.01$

Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, and dummy of Java Island

5.2 Multivariate estimations of current-smoking behaviours on BMI: pooled OLS and FE

Table 7 showed the regression results of pooled OLS and FE models. All estimations used the same number of observations for each group (both sexes, male, female) in any sample (full, current- and non-smoker (CSNS), current- and former-smoker (CSFS), and current-smoker (CS) samples). For the pooled OLS regressions, cluster-robust standard errors were used to control for heteroskedasticity and serial correlations. The pooled OLS results showed a significant negative association (at 1% level) between current smoking and BMI for all samples except for the female group in the CSFS sample. The pooled OLS results regarding the impact of current smoking status on BMI were similar across the full sample and the CSNS sample. For instance, compared with former smokers or non-smokers, current smokers were expected to have a lower BMI by 1.6 points, on average, for both sexes and by around 0.7 to 0.8 point, on average, for males and females, respectively. Recall that the average BMIs for males and females in our dataset were 21.5 and 22.8 respectively. A possible reason why the effect was almost twice as large for both sexes together was that there are natural gender differences in BMI and, for this sample, men were not compared only with men, nor were women only compared with women. Therefore, in discussing the remaining results, I have focused on the male and female sub-samples only.

Comparing current smokers with former smokers (i.e. the CSFS sample), the pooled OLS results showed that the impact of current smoking status on BMI was comparable though slightly smaller than the one in the full sample for males, in which being a current male smoker may be associated with an average of an 0.70-point reduction in BMI. Additionally, the pooled OLS results in the full sample displayed a negative correlation between being a former smoker and BMI for females but not for males. In this case, a former female smoker was expected to have an average of 0.65 point lower BMI than her non-smoker counterparts.

Table 7 Selected coefficients of multivariate estimations on BMI using pooled OLS and FE

Explanatory variables	**Dependent variable: BMI**					
	Full ample					
	Pooled OLS			**FE**		
	Both sexes	**Male**	**Female**	**Both sexes**	**Male**	**Female**
Current smoking status	-1.643***	-0.772***	-0.735***	-0.301***	-0.145**	-0.192
	(0.0400)	(0.0568)	(0.190)	(0.0665)	(0.0722)	(0.174)
Former smoking status	-0.870***	-0.0637	-0.654***	0.00408	0.117	-0.102
	(0.0699)	(0.0793)	(0.237)	(0.0664)	(0.0709)	(0.202)
Observations	72918	33578	38440	72918	33578	38440
	Current-smoker and non-smoker (CSNS) sample					
Current smoking status	-1.641***	-0.768***	-0.740***	-0.249***	-0.0929	-0.0283
	(0.0401)	(0.0570)	(0.190)	(0.0830)	(0.0919)	(0.198)
Observations	69220	30388	38010	69220	30388	38010
The number of cigarettes smoked daily by current smokers	-0.0917***	-0.0242***	-0.0714***	-0.00547*	0.00525*	0.0240
	(0.00283)	(0.00329)	(0.0203)	(0.00317)	(0.00313)	(0.0236)
Observations	68882	30078	37992	68882	30078	37992
	Current-smoker and former-smoker (CSFS) sample					
Current smoking status	-0.722***	-0.696***	0.0797	-0.217***	-0.234***	-0.112
	(0.0688)	(0.0712)	(0.268)	(0.0577)	(0.0599)	(0.215)
Observations	27083	25142	1260	27083	25142	1260
The number of cigarettes smoked daily by current smokers	-0.0140***	-0.00833**	0.0797	0.00226	0.00183	-0.112
	(0.00351)	(0.00354)	(0.268)	(0.00298)	(0.00300)	(0.215)
Observations	26728	24816	1260	26728	24816	1260
	Current-smoker (CS) sample					
The number of cigarettes smoked daily by current smokers	0.00843**	0.0142***	-0.0324	0.0150***	0.0144***	0.0581
	(0.00397)	(0.00400)	(0.0280)	(0.00327)	(0.00330)	(0.0408)
Observations	23047	21642	812	23047	21642	812

Cluster-robust standard errors in parentheses
$^{*}P < 0.1$, $^{**}P < 0.05$, $^{***}P < 0.01$

Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption

In the CSNS and CSFS samples, the pooled OLS results also revealed that the number of cigarettes consumed daily by current smokers had a negative impact on both males' and females' BMI. The impact of the intensity of smoking on BMI seemed smaller than the impact of being a current smoker, in which one more cigarette stick consumed per day may reduce BMI by less than 0.1 point across all samples, while current smoking status was associated with an average of about 0.7 to 1.6 points

decrease in BMI. In the CSNS sample, the pooled OLS outcomes regarding smoking intensity could imply that compared with non-smokers, smokers' BMI was likely to reduce, on average, by 0.02 point for males and 0.07 point for females, for each additional cigarette stick consumed per day. When comparing current smokers and former smokers (the CSFS sample), the results showed insignificant effects of smoking intensity on females' BMI and much lower effects on males' BMI compared with the CSNS sample. There was a significant but a small effect size for males in which every additional cigarette stick consumed per day may decrease BMI by 0.008 point, on average, compared with those who already quit smoking. Among current smokers (the CS sample), the number of cigarettes consumed daily was positively correlated with BMI for males. Every additional cigarette stick consumed daily increased BMI by 0.01 point, on average. Again, smoking intensity was not significant among female current smokers.

Even though most of the outcomes from pooled OLS estimations depicted significant correlations between smoking and BMI, they could not be used to draw causal inferences due to potential endogeneity problems from unobserved heterogeneity, time-variant errors, and reverse causality. By using methods without addressing endogeneity problems, some previous studies found similar results. For instance, previous studies discovered that the average number of cigarettes consumed per day was negatively correlated with BMI when comparing current smokers and non-smokers, but was positively associated with BMI among current smokers only (Bamia et al., 2004; Dare et al., 2015). Other studies using similar methods also obtained equivalent results that the current smoking status—with non-smokers as the reference group—(Plurphanswat & Rodu, 2014; Watanabe et al., 2016) and the number of cigarettes consumed by current smokers, compared with all non-current smokers, was negatively correlated with BMI (Fang et al., 2009; Sohn, 2015).

The results from the FE estimations presented in Table 7 showed the impact of current-smoking behaviours on BMI after correcting the endogeneity problems caused by unobserved heterogeneity or time-invariant factors, such as adverse childhood experience or time preference. Compared with the pooled OLS results, the coefficients from the fixed-effect models had different significance for different groups in various samples, and the magnitude tended to be smaller than that estimated using the pooled OLS method, indicating the pooled OLS method may overestimate the impacts of

current-smoking behaviours on BMI. Fixed-effect estimations in the full sample depicted an adverse effect of the current-smoking status on males' BMI (but was insignificant for females') with a smaller coefficient than the pooled OLS one, and showed that the former smoking status was no longer significant. Notably, for men, current smoking status was associated with a reduction in BMI by only 0.15 points, on average, under FE methods compared with a reduction in BMI by 0.77 points, on average, under pooled OLS. Regarding the effects of current smoking status–in other samples, being a current smoker did not significantly affect males and females in the CSNS sample and females in the CSFS sample. In the CSFS sample, the effects of being a male current smoker may be associated with only a 0.23-point reduction in BMI compared with a 0.70-point reduction in the pooled OLS. There was also an evidence (at the 10% level of significance) of a relatively small positive impact of the number of cigarettes consumed daily on BMI in the male group of the CSNS sample, but with a considerably smaller impact than the pooled OLS results. Furthermore, the intensity of smoking of current smokers was insignificant for both males and females compared with former smokers, while it was positively correlated with BMI among male current smokers but only with a similar magnitude to the pooled OLS result.

The results from FE estimations still could not entirely remove endogeneity problems in the model, particularly those coming from unobserved time-variant errors and reverse causality. Therefore, this study used fixed effects with instrumental variables (FEIV) to tackle endogeneity issues thoroughly. The model of FEIV is equivalent to the second-stage regression of current-smoking behaviours, in which the endogenous variables were estimated with instrumental variables in the first-stage regression in section 5.1. The analysis of the FEIV models will be discussed further in the following section.

5.3 FEIV: second-stage regression of current-smoking behaviours on BMI

This section analysed the impacts of current smoking status and the number of daily cigarettes smoked on BMI using FEIV to fully remove the endogeneity issues. The regressions were estimated with and without per capita oil and sugar consumption to see whether the effect of current-smoking indicators was partly because of the

influence of cigarette smoking on basic food consumptions. In this study, the observations under the same group and sample (e.g. male in the full sample) were set as identical in the analysis with and without food consumptions to avoid the change in current-smoking coefficients owing to sample difference.

5.3.1 Second-stage regression: the impact of current smoking status on BMI

Table 8 shows the results for the critical variables regarding the causal effect of current smoking status on BMI estimated with FEIV for the full sample, CSNS sample, and CSFS sample (see Appendix E for complete regression results). The regression outcomes in the full sample enabled me to explore the causal inference of current smoking status compared with both former smokers and non-smokers, as well as to analyse the impact of former smoking status with current smokers and non-smokers as reference groups. In contrast with the results from the models with pooled OLS and FE, the outcomes of FEIV revealed that current smoking status did not have significant impacts on BMI for males and females in any of the three samples after completely removing endogeneity problems from the estimations. It can be inferred that the results from the pooled OLS and FE models, which still suffered from the endogeneity issues, may provide a misleading idea of significant negative correlation between current smoking status and BMI compared with non-smokers and former smokers.

According to the effect of former smoking in the full sample, the outcomes showed that being a former smoker did not significantly alter BMI (higher or lower) compared to being a non-smoker. The addition of per capita oil and sugar consumption in the model did not change the significance of former-smoking status. Further analysis of the impact of former-smoking behaviours can be found in section 5.4. In the CSNS sample, the results implied that being a current smoker, compared with being a non-smoker, did not have any influence on BMI irrespective of the inclusion of variables to control oil and sugar consumption. Similar results were also observed in the CSFS sample: being a current smoker did not significantly affect BMI when compared with being a former smoker. Overall, the inclusion of sugar and oil consumption in the model

tended to change the magnitude of the coefficients but did not change their significance for current smoking status across all samples.

Table 8 The impacts of current smoking status on BMI using FEIV

Explanatory variables	**Dependent variable: BMI**					
	Full sample					
	Both sexes		**Male**		**Female**	
Current smoking status	-7.06	-4.321	3.48	4.7	-81.52	-64.03
	(4.615)	(4.274)	(2.961)	(3.145)	(75.50)	(55.33)
Former smoking status	-4.879	-2.901	2.75	3.636	-51.26	-40.25
	(3.338)	(3.090)	(2.151)	(2.285)	(47.59)	(34.88)
Log effective per capita oil consumption		0.0775***		0.0681***		0.0958
		(0.0169)		(0.0233)		(0.0669)
Log effective per capita sugar consumption		0.0475***		0.0554**		0.0865
		(0.0168)		(0.0239)		(0.0915)
Observations	72918	72918	33578	33578	38440	38440
	Current-smoker and non-smoker (CSNS) sample					
	Both sexes		**Male**		**Female**	
Current smoking status	-9.094	-5.945	4.256	5.537	-117.3	-93.04
	(5.789)	(5.291)	(3.300)	(3.498)	(128.9)	(94.51)
Log effective per capita oil consumption		0.0774***		0.0802***		0.0799
		(0.0191)		(0.0258)		(0.0873)
Log effective per capita sugar consumption		0.0432**		0.0368		0.102
		(0.0192)		(0.0267)		(0.123)
Observations	69220	69220	30388	30388	38010	38010
	Current-smoker and former-smoker (CSFS) sample					
	Both sexes		**Male**		**Female**	
Current smoking status	0.188	0.708	0.157	0.657	-0.0149	0.997
	(1.138)	(1.150)	(1.241)	(1.260)	(3.091)	(2.765)
Log effective per capita oil consumption		0.0448**		0.0379*		0.165
		(0.0217)		(0.0228)		(0.116)
Log effective per capita sugar consumption		0.0733***		0.0717***		0.0648
		(0.0227)		(0.0241)		(0.141)
Observations	27083	27083	25142	25142	1260	1260

Cluster-robust standard errors in parentheses

$^{*}P < 0.1$, $^{*\mathrm{P}}\ pP < 0.05$, $^{***}P < 0.01$

Endogenous variable: Current smoking status
Instrumental variable of current smoking status: Percentage of current smokers in the community
Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption

5.3.2 Second-stage regression: the impact of the number of cigarettes smoked daily by current smokers on BMI

In the study, I also included the number of cigarettes smoked daily by current smokers to measure the effects of smoking intensity of current smokers on BMI. As seen in Table 9, after having controlled the endogeneity problems, I found that the number of cigarettes consumed had no significant causal effects in almost all groups and samples, except for the one with combined male and female observations (both sexes) in the CSNS sample. The inclusion of per capita oil and sugar consumption slightly changed the magnitude (but not the significance) of the coefficient on number of cigarettes consumed by current smokers.

The positive results in the combined group of men and women in the CSNS sample inferred that compared with non-smokers, one more cigarette stick consumed daily by combined male and female current smokers could potentially increase their BMI by around 0.73 points, on average. However, it seemed that the result mostly comes from the variations in the female group as the coefficient for the male sub-sample was insignificant. Provided that the first-stage regressions of the smoking intensity for females in all samples did not have a valid instrumental variable, the result may not be considered as a key finding of the study.

The estimations of smoking intensity with pooled OLS and FE models may report incorrect inference due to the endogeneity issues. After correcting for the issues, what was presented in Table 7 (i.e. a negative effect of the number of cigarettes consumed daily on smokers compared with non-smokers and former smokers, and its positive effect among current smokers) seemed not to be the case in Indonesia. The results providing causal inference in this report supported the findings by Chen et al. (2007) which compared OLS results and causal inference from simultaneous equation models. In line with my findings, the results by Chen et al. (2007) showed that the average number of cigarettes consumed per day may have significant negative correlation with BMI under OLS models, but it had no significant impacts after completely controlling for endogeneity.

Table 9 The impacts of the number of cigarettes smoked by current smokers on BMI using FEIV

Explanatory variables	**Dependent variable: BMI**			
	Current-smoker and non-smoker (CSNS) sample			
	Both sexes		**Male**	
The number of daily cigarettes smoked by current smokers	0.725***	0.690***	-0.101	-0.155
	(0.222)	(0.235)	(0.0851)	(0.0980)
Log effective per capita oil consumption		0.0607**		0.0814***
		(0.0258)		(0.0256)
Log effective per capita sugar consumption		-0.0196		0.0829***
		(0.0311)		(0.0276)
Observations	68882	68882	30078	30078
	Current-smoker and former-smoker (CSFS) sample			
	Both Sexes		**Male**	
The number of daily cigarettes smoked by current smokers	-0.031	-0.0924	-0.0562	-0.113
	(0.109)	(0.133)	(0.107)	(0.131)
Log effective per capita oil consumption		0.0669*		0.0681*
		(0.0356)		(0.0391)
Log effective per capita sugar consumption		0.0889***		0.0849***
		(0.0309)		(0.0283)
Observations	27083	27083	25142	25142
	Current-smoker (CS) Sample			
	Both sexes		**Male**	
The number of daily cigarettes smoked by current smokers	-0.015	-0.0443	-0.0243	-0.0534
	(0.0571)	(0.0634)	(0.0581)	(0.0646)
Log effective per capita oil consumption		0.0588**		0.0634**
		(0.0253)		(0.0263)
Log effective per capita sugar consumption		0.0749***		0.0689**
		(0.0280)		(0.0277)
Observations	23047	23047	21642	21642

Cluster-robust standard errors in parentheses
$^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$
Endogenous variable: current smoking status
Instrumental variable of current smoking status: percentage of current smokers in the community
Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption

5.4 Multivariate estimations of former-smoking behaviours on BMI

In this study, not only did I estimate the effects of current-smoking behaviours on body measurement, but I also explored the role of former-smoking behaviours on BMI compared with non-smokers as depicted in Table 10. All of the former-smoking variables were estimated separately[8] using a FE model without instrumental variables since there seemed no valid instruments for this variable (as explained in section 4.3). Therefore, it should be noted that the results presented in this section were limited to non-causal inference due to the potential existence of the endogeneity problems from time-variant errors and reverse causality. There were four former-smoking-related explanatory variables estimated in the regressions: 1) former smoking status (dummy), 2) the number of cigarettes consumed per day before quitting smoking, 3) duration of smoking (in years) before giving up smoking, and 4) length of time since quitting smoking (the gap years from quitting smoking until the time of interview).

In the analysis that was not modified by sex (including both sexes), the results showed significant negative impacts of former smoking status, duration of smoking before quitting smoking, and length of time since quitting smoking, but no observed significant effects of smoking intensity before quitting smoking. It inferred that, irrespective of sex, former smokers tended to have lower BMI than their non-smoker counterparts. Additionally, a longer duration of smoking before quitting and a longer period of smoking absence tended to decrease BMI, but only with 10% level of significance. However, when the results were separated by gender, the significant effects of former-smoking behaviours vanished. It indicated that the pattern of BMI and former-smoking behaviours for men and women were different (Plurphanswat & Rodu, 2014; Watanabe et al., 2016) and that comparing them may result in an incorrect conclusion. Thus, based on the results modified by gender, I concluded that former smoking status, the number of cigarettes smoked per day before quitting smoking, duration of smoking before quitting smoking, and length of time since quitting smoking did not have any significant effects on BMI in men and women.

As opposed to the results of my study, many earlier studies found that smoking cessation was associated with weight gain (Filozof et al., 2004; Kasteridis & Yen, 2012; Pieroni & Salmasi, 2016) as well as higher BMI (Z. Chen et al., 2007; Courtemanche

[8] One former-smoking related variable in one regression.

et al., 2018; Piirtola et al., 2018; Watanabe et al., 2016) compared with never smoking. However, several previous studies also observed similar findings to mine—that the BMI of former smokers was not significantly different from those who never smoked (Guerra et al., 2015; U. John et al., 2005; Køster-Rasmussen et al., 2015). Regarding smoking duration before quitting smoking, my findings were supported by Watanabe et al. (2016)—establishing that the smoking duration of former smokers did not have any effects on BMI. Additionally, Watanabe et al. (2016) also confirmed that the length of time since quitting smoking was insignificantly related to BMI.

Table 10 The effects of former-smoking behaviours on BMI using FE

	Dependent variable: BMI		
	Both sexes	**Male**	**Female**
Former smoking status	-0.193**	-0.109	-0.388
	(0.0853)	(0.0880)	(0.275)
Observations	49533	11626	37610
The number of daily cigarettes smoked by former smokers before giving up smoking	-0.00469	-0.00199	-0.00622
	(0.00564)	(0.00562)	(0.0275)
Observations	49533	11626	37610
Duration of smoking of former smokers before giving up smoking	-0.00755*	-0.00508	-0.00855
	(0.00440)	(0.00467)	(0.0127)
Observations	47979	10345	37362
Length of time since quitting smoking	-0.00727*	-0.00301	-0.0205
	(0.00413)	(0.00436)	(0.0155)
Observations	48048	10403	37371

Cluster-robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption

5.5 The Impact of other covariates on BMI

In this report, I also included a set of additional explanatory variables to help explain the variation of BMI. The effects of the incorporated regressors can be seen in Appendix E, F, and G for the regressions of current-smoking status, the number of cigarettes consumed daily, and former-smoking behaviours, respectively. Most coefficient signs of the variables were anticipated and followed the directions expected in the literature. In terms of the individual characteristics, age had a significant positive impact on BMI in almost all regressions, but the impact of age seemed to decrease over time (significant negative coefficient of age square). The outcome followed the theoretical arguments in the literature that the elderly were more likely to be less active and that they tend to have a surplus energy balance[9] than people in other age groups (Jura & Kozak, 2016) and previous evidence of the negative effects of age square on BMI (Z. Chen et al., 2007; Plurphanswat & Rodu, 2014; Sohn, 2015). Marital status and total working hours were mostly insignificant in the regressions but found to have significant positive effects in the estimations of former-smoking behaviours. The tendency of married people to have higher BMI than their unmarried counterparts was expected, based on findings by other researchers (Guerra et al., 2015; Liao et al., 2018; Tzotzas et al., 2010). However, the lack of a relationship between the two was also possible as the dynamic of marital transition may explain BMI and body weight better than the naive measure of marital status (Dinour, Leung, Tripicchio, Khan, & Yeh, 2012; Sobal & Hanson, 2011). The insignificant effect of total working hours can be deduced by the effect of physical activities embedded in a job. It is possible that the positive impacts of non-strenuous and white-collar jobs may be outweighed by the negative influences of strenuous and blue-collar jobs on body weight (Abramowitz, 2016; Dang, Maitra, & Menon, 2019). I also estimated that the years of schooling was positively related to BMI in some regressions, which was in line with the finding in a similar paper by Baum (2009).

Among the included household characteristics, location of households in either urban areas or a more-developed island (Java Island) did not significantly influence BMI. The effective household per capita income, as well as oil and sugar consumptions, were observed as positively correlated with BMI in most of the regression results in

[9] Surplus energy balance may take place when calorie intake is greater than the calorie burnt for energy.

this study. The authors of another study, examining the contributing factors of overweight/obesity in Indonesia, also observed positive effects of income on weight status (Rachmi, Li, & Alison Baur, 2017). As stated by Cawley (2015), the positive relationship between income and body weight was possible if people value sedentary leisure and food, including unhealthy foods, higher than the value of future health and body image. However, there is a need for further evaluation of the relationship as the impacts of income may have a dissimilar direction at a different level of income and tail of BMI distribution (Jolliffe, 2011; Lakdawalla & Philipson, 2009; Tafreschi, 2015). In terms of consumption, oil and sugar per capita intake did not change the significance of current smoking on BMI. The positive effect of both indicators seemed to be commonly recognised and was also confirmed by previous related studies (Bentley, Ruck, & Fouts, 2019; Qi et al., 2014).

5.6 Sensitivity analysis

Besides the regressions in the result section, I also conducted several robustness checks to measure the sensitivity of the results to various modifications of the models. Because of space consideration, I put the results of sensitivity analyses in Appendix H. The sensitivity analyses were conducted as follows. First, I ran a regression with balanced panel data, which implied that only people who participated for the whole period were included in the estimation (see Appendix H.1). This procedure was meant to see whether the results were based on the problems of attrition. After removing all observations that did not appear consistently in all five waves, the data was balanced and included 26,135 observations (5,227 people observed in each wave for all five waves) consisting of 15,945 males and 10,190 females[10]. The regressions revealed no significant correlations between current smoking status, the number of cigarettes consumed, and former smoking variables and BMI across all samples for both males and females, except for the number of cigarettes consumed before quitting smoking in the female sample, which had a positive significant impact with BMI. Even though the

[10] To be noted that the number of observations here matched the data before running any regressions, while the appendix H.1 shows the number of observations in each regression included in the sensitivity analysis. A strongly balanced data by including only observations appeared in all five waves *and* with non-missing values in all variables was not feasible due to missing values in the variables.

result might indicate that a former smoking habit was positively correlated with BMI, the result showed only a small effect at 5% level of significance.

Second, I divided the samples into several age groups (see Appendix H.2). As indicated by previous researchers, the impacts of current and former smoking behaviours may differ according to the age group that one belongs to (Courtemanche et al., 2018; Dare et al., 2015; Kasteridis & Yen, 2012). All the previous researchers used the cut-off of below 40–45 years old as the lowest age cohort, while the groups above 40–45 years old varied from one study to another. I adopted the age groups in the literature, separating the samples into three age categories: 15–40 years, 41–60 years, and above 60 years. In order to present an uncomplicated explanation of the robustness checks, only the CSNS sample was used to analyse the current smoking status and smoking intensity and only the FSNS sample was used to examine former smoking status. The estimations showed no coefficient was significant at least at the 5% confidence level, except for former smoking status among elderly females (above 60 years old). It was depicted that elderly female former smokers tend to have lower BMI of 0.79 point, on average, compared with their non-smoker counterparts. Based on the result, I extended the estimations to analyse the impacts of other former smoking behaviours (see Appendix H.3). Former smoking status, duration of smoking before quitting smoking, and the period of smoking absence had significant negative coefficients at 1% level of confidence and were associated with the reduction of mean BMI by 0.79, 0.02, and 0.07 points, respectively. However, smoking intensity before quitting smoking was observed as having negative effects but with only a 10% level of significance. The result is a rather unexpected outcome. At this stage, I had not intended to explore the impacts of former smoking behaviours for a particular age range, but this may indicate a unique relationship between former smoking and body weight which could be useful for further studies.

Third, several regressions were conducted only for 1) people who were at least once pre-obese or obese across the whole period, and 2) those who experienced pre-obese or obesity at the time of interview. The former group meant that the estimations include people who were not pre-obese or obese at the time of the interview, but had experienced pre-obesity or obesity at least once across the five waves, while the latter group denoted that the regressions were conducted only for people who were pre-obese or obese at the time of the interview. The inclusion of the observations who were once

pre-obese or obese aimed to allow for fluctuation of body weight which may be influenced by smoking. Similar to previous robustness checks, the regressions were run only for the current and non-smoker (CSNS) sample to examine current smoking status and smoking intensity, and run only for the former and non-smoker (FSNS) sample to evaluate former smoking status. The analysis was based on several earlier studies examining that smoking may affect body weight variously across different weight status (Fang et al., 2009; Plurphanswat & Rodu, 2014; Wang, 2015; Wehby et al., 2012). The weight categories were adopted from Asian-standard weight status as explained in section 3.2.1. The estimations revealed that there were no observed significant effects of current smoking status, smoking intensity, and former smoking status for people who were once pre-obese or obese and for those who were pre-obese or obese.

6 Discussion

The present study was designed to determine the causal effect of current smoking status and smoking intensity as well as the impact of former smoking on BMI in Indonesia. After removing the problem of endogeneity by using FEIV methods, I found that—when comparing smokers with all non-current smokers, non-smokers, and former smokers—current smoking status did not have any significant causal effects on BMI in Indonesia for both males and females. Similarly, I did not detect any evidence for significant causal impacts on BMI in men from the number of cigarettes consumed per day. Another important finding was that there were no observed significant non-causal effects of several measures of quitting smoking on BMI for both males and females.

The outcomes in my study were contrary to most previous studies, which have suggested that current smoking may reduce body weight, and a few studies indicating the opposite effect. Regarding current smoking, the results were not supportive of the idea of the *slimming effects* of smoking and that increased weight was an unintended consequence of smoking (Chou et al., 2004; Rashad & Grossman, 2004). Mainly, my findings were not in line with previous studies indicating the negative causal influence of current smoking status and smoking intensity on body weight (Courtemanche et al., 2018; Fang et al., 2009; Sohn, 2015; Wang, 2015; Wehby et al., 2012). Neither did my results agree with some earlier studies reporting the positive causal effect of current smoking on body weight (Gruber & Frakes, 2006; Wehby & Courtemanche, 2012). In contrast with the results by Bamia et al. (2004) and Dare et al. (2015), I also found that, among current smokers, the number of cigarettes consumed per day may not positively affect BMI.

The results corroborate the ideas of Chen et al. (2007) as they found that smoking might have adverse effects under OLS but revealed insignificant causal effects of the number of cigarettes consumed per day on BMI after controlling for the endogeneity. My research was also in line with research by Nonnemaker et al. (2009) suggesting that the effects of cigarette price, which may reduce the smoking intensity or decrease smoking prevalence, was insignificantly correlated with body weight. It was likely that the negative impact of smoking on body weight reported by most researchers—using OLS or other methods that did not fully remove the endogeneity problems—might be a biased result due to misspecification of the models (Z. Chen et al., 2007). A similar

study in Indonesia by Sohn (2015) was no exception. His findings of the inverse correlation of smoking on BMI might still suffer the remaining endogeneity problems from reverse causality and time-variant errors and was likely to provide a fallacious inverse relationship between smoking status and BMI in Indonesia. As my research revealed, the use of FE methods, as used in Sohn's study, showed negative effects of current smoking status and smoking intensity in men and women smokers compared with non-smokers and former smokers. However, the effects disappeared after the rest of the endogeneity problems were removed by instrumental variables, resulting in insignificant effects of current smoking on BMI in Indonesia.

Contrary to expectations, I did not find significant causal effects of current smoking status and smoking intensity on body weight. A possible explanation for this might be that both negative and positive impacts of smoking may exist so that the inverse effects outweigh each other. As an illustration, on the one hand, smoking may induce lower appetite resulting in lower caloric intake (Audrain-Mcgovern & Benowitz, 2011; H. Chen et al., 2012), but on the other hand it may also promote sedentary behaviours and lower exercise capacity (Kaufman et al., 2012; Rodriguez et al., 2010), and increase insulin resistance which may increase body weight (Chiolero et al., 2008; Schindler et al., 2006). Another possible explanation could be the differences in research populations. For example, the negative causal effect of smoking may be the case in China, as two studies observed that smoking intensity tended to decrease BMI as well as the likelihood of being obese in Chinese people (Fang et al., 2009; Wang, 2015). Other examples of similar results were from the United States, which is a developed country (Courtemanche et al., 2018), and a small sample of mothers in Norway (Wehby et al., 2012). However, these findings from other countries may not apply to Indonesia. As there were many different possible mechanisms linking smoking to body weight, the obvious consequences of smoking on body weight may differ from one country to another due to different circumstances and factors, mediating smoking and body weight (Sohn, 2015). For example, a society with poverty and food insecurity that induces unhealthy eating may not support the *slimming effect* of smoking suppressing appetite as the effect may be outweighed by the higher caloric intake from unhealthy eating. This reason could be the case in Indonesia as most Indonesian smokers were in the lower-income groups, experiencing poverty and vulnerability (Amalia et al., 2019; Dartanto, Nurhasana, Thabrany, Moeis, & Satrya, 2018).

In this study, I have been unable to demonstrate that former smoking may be positively correlated with BMI as suggested by previous studies analysing causal effects of former smoking (Courtemanche et al., 2018; Kasteridis & Yen, 2012; Pieroni & Salmasi, 2016). However, the results should be interpreted with caution as they did not reveal the causal impacts of quitting smoking. It is important to consider that possible bias might still exist due to the remaining endogeneity problems from unobserved time-variant factors and reverse causality. Even though most previous studies observed positive impacts on BMI of former-smoking behaviours, a number of papers found that former smokers could have similar weight as non-smokers (Pisinger et al., 2017), depending on their BMI status and age (Bossé, Garvey, & Costa, 1980; Pisinger et al., 2017). It seems that the emphasis of previous studies on the higher BMI of former smokers might contribute to the misleading deduction in the literature and society that increased weight is one of the inevitable consequences of quitting smoking (Pisinger et al., 2017). It is plausible that the positive effect of smoking cessation on BMI may only be a prompt effect which may disappear as the period of smoking abstinence increases (Froom et al., 1999; Kamaura, Fujii, Mizushima, & Tochikubo, 2011; Mizoue, Ueda, Tokui, Hino, & Yoshimura, 1998; Munafò, Tilling, & Ben-Shlomo, 2009; Reas, Nygård, & Sørensen, 2009; Scherr et al., 2015). After around four to six years after quitting smoking, former smokers were expected to have a similar BMI to non-smokers (Froom et al., 1999; Kamaura et al., 2011; Reas et al., 2009). This was one of the potential reasons for no correlation between quitting smoking and BMI in Indonesia, provided that the median duration of smoking abstinence in IFLS data was six years for males and five years for females. Moreover, most positive effects of quitting smoking or positive effects of cigarettes costs on body weight were observed in the United States (Baum, 2009; Chou et al., 2004; Courtemanche et al., 2018; Kasteridis & Yen, 2012; Rashad & Grossman, 2004). One possible explanation is that in the United States, people might substitute cigarettes with unhealthy foods as there was massive availability of fast food restaurants in the country (Cawley, 2015).

The insignificant role of smoking on body weight suggests that some other factors were responsible for the rise of obesity in Indonesia. The results from other covariates in this study indicated that ageing, education, income, and food consumptions might predict BMI in Indonesia. Other indicators—such as the use of modern technology in households that promotes sedentary behaviours, exercise and physical activities, the

availability of fast-food restaurants and convenience stores in the neighbourhood, the unemployment rate, and social perceptions about body weight—may possibly play more significant roles in body weight (Cawley, 2015; Lakdawalla & Philipson, 2009).

7 Conclusion

Previous studies evaluating the impacts of smoking on body weight have observed inconclusive results on whether smoking affected body weight positively, negatively, or insignificantly. Analysing panel data from the IFLS 1993–2014, I determined the causal effect of current smoking and the role of former smoking on body weight in Indonesia in males and females separately. By entirely removing the endogeneity problems of current smoking, the study revealed that current smoking status and the number of cigarettes smoked per day did not have any significant causal relationship with BMI. Furthermore, even though the analysis of former smoking was limited to non-causal inference, the findings suggested that former smoking status, smoking intensity and duration of smoking before quitting smoking, and length of time since giving up smoking did not have any significant association with BMI. Overall, based on the findings in this study, it can be assumed that smoking should not be used as a means of controlling weight as there were no observed causal effects of smoking on body weight. Moreover, the findings suggested that one should not be daunted from quitting smoking due to weight concern since there were no significant positive effects on weight from giving up smoking.

Like any other research, this study has a number of shortcomings. First, I used instrumental variables from community-smoking indicators which may be sensitive to sampling methods. I encourage the use of more convincing instruments—such as gene and DNA used in one of the previous studies, or other methods to eradicate endogeneity problems—in further studies. Second, in this study I did not aim to measure heterogeneous effects of smoking on different measures of body weight or BMI of various groups of people. Since previous studies indicated dissimilar magnitude and significance of smoking across different BMI distribution and weight status, future studies may want to extend this study to explore various impacts of smoking on different levels of weight measures. Additionally, further research needs to examine more closely the links between smoking and body weight in different groups of people. Third, I only used one measure of body weight and fatness (BMI), and there are other measures such as waist-to-hip ratio or waist circumference which, though not employed in this study, may arguably be better alternative measures of fatness. Fourth, I did not explore the mechanism of how smoking could affect body weight. Further research

exploring the issue would be of great help in understanding the reasons behind causal impacts of smoking. Fifth, as this study only analysed people aged 15 years and above, the phenomenon of smoking among children and teenagers in Indonesia was beyond the scope of this study. This research area would be a fruitful area for further work. Sixth, the results of former smoking behaviours may be somewhat limited by the fact that it may still suffer from the endogeneity problems from time-variant errors and reverse causality. Considerably more work will need to be done to determine the causal impacts of quitting smoking on body weight by implementing valid instruments or other endogeneity-removing methods.

Notwithstanding the limitations of the study, my research offers valuable insights into the literature regarding the causal effects of current smoking and the possible role of former smoking on body weight in developing countries and, specifically, Indonesia. Moreover, my research appears to be the first study in Indonesia to analyse the causal impacts of current smoking by fully eradicating the endogeneity problems, and is a pioneering study exploring the association of former-smoking behaviours and BMI. The findings of this study also provide insights into the issue of related policies in Indonesia. On the one hand, any anti-smoking policies in Indonesia should not be discouraged due to the fear of an increase in the obesity rate as an unintended consequence of quitting smoking, as this study did not observe any significant *slimming effect* of smoking. Given the serious smoking-related illnesses and healthcare costs, the erroneous inference that potentially persuades the general population to smoke as an obesity-prevention or weight-control method may exacerbate the future financial burden and the catastrophic health problems of the country. On the other hand, as there was no observed positive effect of smoking on body weight, I cannot suggest that the policy to combat smoking may have the double effect of also reducing obesity levels. Instead, even though smoking did not have any causal impact on body weight, continued efforts are needed to combat both issues simultaneously as smoking and obesity may cause the decline in economic development through lower human capital and escalated healthcare costs.

References

Abramowitz, J. (2016). The connection between working hours and body mass index in the U.S.: a time use analysis. *Review of Economics of the Household*, *14*(1), 131–154.

Aizawa, T., & Helble, M. (2017). Socioeconomic inequality in excessive body weight in Indonesia. *Economics and Human Biology*, *27*, 315–327.

Alkerwi, A., Baydarlioglu, B., Sauvageot, N., Stranges, S., Lemmens, P., Shivappa, N., & Hébert, J. R. (2017). Smoking status is inversely associated with overall diet quality: Findings from the ORISCAV-LUX study. *Clinical Nutrition*, *36*, 1275–1282. https://doi.org/10.1016/j.clnu.2016.08.013

Amalia, B., Cadogan, S. L., Prabandari, Y. S., & Filippidis, F. T. (2019). Socio-demographic inequalities in cigarette smoking in Indonesia, 2007 to 2014. *Preventive Medicine*, *123*, 27–33. https://doi.org/10.1016/j.ypmed.2019.02.025

Amir, S. (1996). Smoking cessation and health: A comment. *Journal of Health Economics*, *15*, 751–754. Retrieved from http://eprints.whiterose.ac.uk/56746/

Au, N., Hauck, K., & Hollingsworth, B. (2013). The relationship between smoking, quitting smoking and obesity in Australia: A seemingly unrelated probit approach. *Applied Economics*, *45*(16), 2191–2199. https://doi.org/10.1080/00036846.2012.657353

Audrain-Mcgovern, J., & Benowitz, N. L. (2011). Cigarette smoking, nicotine, and body weight. *Clinical Pharmacology and Therapeutics*, *90*(1), 164–168. https://doi.org/10.1038/clpt.2011.105

Bamia, C., Trichopoulou, A., Lenas, D., & Trichopoulos, D. (2004). Tobacco smoking in relation to body fat mass and distribution in a general population sample. *International Journal of Obesity*, *28*, 1091–1096. https://doi.org/10.1038/sj.ijo.0802697

Barber, S., Adioetomo, S. M., Ahsan, A., & Setyonaluri, D. (2008). *Tobacco economics in Indonesia. International Union Against Tuberculosis and Lung Disease.* Retrieved from http://www.worldlungfoundation.org/ht/a/GetDocumentAction/i/6567

Baum, C. L. (2009). The effects of cigarette costs on BMI and obesity. *Health Economics*, *18*, 3–18.

Bentley, R. A., Ruck, D. J., & Fouts, H. N. (2019). U.S. obesity as delayed effect of excess sugar. *Economics and Human Biology*, 1–7. https://doi.org/10.1016/j.ehb.2019.100818

Berk-Clark, C. van den, Secrest, S., Walls, J., Hallberg, E., Lustman, P. J., Schneider, F. D., & Scherrer, J. F. (2018). Association Between Posttraumatic Stress Disorder and Lack of Exercise , Poor Diet , Obesity, and Co-Occuring Smoking : A Systematic Review and Meta-Analysis. *Health Psychology*, *37*(5), 407–416.

Blaine, B. (2008). Does depression cause obesity?: A meta-analysis of longitudinal studies of depression and weight control. *Journal of Health Psychology*, *13*(8), 1190–1197. https://doi.org/10.1177/1359105308095977

Bossé, R., Garvey, A. J., & Costa, P. T. (1980). Predictors of weight change following smoking cessation. *Substance Use and Misuse*, *15*(7), 969–991. https://doi.org/10.3109/10826088009040072

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*(430), 443–450. https://doi.org/10.1080/01621459.1995.10476536

Bush, T., Hsu, C., Levine, M. D., Magnusson, B., & Miles, L. (2014). Weight gain and smoking: Perceptions and experiences of obese quitline participants. *BMC Public Health*, *14*(1), 1–9. https://doi.org/10.1186/1471-2458-14-1229

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. *The British Journal of Psychiatry* (Vol. 112). Cambridge: Cambridge University Press.

Carreras-Torres, R., Johansson, M., Haycock, P. C., Relton, C. L., Davey Smith, G., Brennan, P., & Martin, R. M. (2018). Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. *BMJ*. https://doi.org/10.1136/bmj.k1767

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, *43*, 244–268.

Cawley, J., Markowitz, S., & Tauras, J. (2004). Lighting up and slimming down: The effects of body weight and cigarette prices on adolescent smoking initiation. *Journal of Health Economics*, *23*, 293–311.

Cawley, J., Markowitz, S., & Tauras, J. (2006). Obesity, Cigarette Prices, Youth

Access Laws, and Adolescent Smoking Initiation. *Eastern Economic Journal*, *32*(1), 149–170.

Cawley, J., & Ruhm, C. J. (2011). The Economics of Risky Health Behaviors. In *Handbook of Health Economics* (Vol. 2, pp. 95–199). Elsevier B.V. https://doi.org/10.1016/B978-0-444-53592-4.00003-7

Chen, H., Saad, S., Sandow, S. L., & Bertrand, P. P. (2012). Cigarette smoking and brain regulation of energy homeostasis. *Frontiers in Pharmacology*, *3*(147), 1–8. https://doi.org/10.3389/fphar.2012.00147

Chen, Z., Yen, S. T., & Eastwood, D. B. (2007). Does smoking have a causal effect on weight reduction? *Journal of Family and Economic Issues*, *28*(1), 49–67.

Chiolero, A., Faeh, D., Paccaud, F., & Cornuz, J. (2008). Consequences of smoking for body weight, body fat distribution, and insulin resistance. *American Journal of Clinical Nutrition*, *87*, 801–809.

Chiolero, A., Jacot-Sadowski, I., Faeh, D., Paccaud, F., & Cornuz, J. (2007). Association of cigarettes smoked daily with obesity in a general adult population. *Obesity*, *15*(5), 1311–1318. https://doi.org/10.1038/oby.2007.153

Chou, S.-Y., Grossman, M., & Saffer, H. (2004). An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics*, *23*, 565–587.

Clair, C., Chiolero, A., Faeh, D., Cornuz, J., Marques-Vidal, P., Paccaud, F., ... Vollenweider, P. (2011). Dose-dependent positive association between cigarette smoking, abdominal obesity and body fat: Cross-sectional data from a population-based survey. *BMC Public Health*, *11*(23), 1–10. https://doi.org/10.1186/1471-2458-11-23

Costa-Font, J., & Gil, J. (2005). Obesity and the incidence of chronic diseases in Spain: A seemingly unrelated probit approach. *Economics and Human Biology*, *3*, 188–214. https://doi.org/10.1016/j.ehb.2005.05.004

Courtemanche, C. (2009). Rising cigarette prices and rising obesity: Coincidence or unintended consequence? *Journal of Health Economics*, *28*(4), 781–798.

Courtemanche, C., Tchernis, R., & Ukert, B. (2018). The effect of smoking on obesity: Evidence from a randomized trial. *Journal of Health Economics*, *57*, 31–44. https://doi.org/10.1016/j.jhealeco.2017.10.006

Dang, A., Maitra, P., & Menon, N. (2019). Labor market engagement and the body

mass index of working adults: Evidence from India. *Economics and Human Biology*, *33*, 58–77. https://doi.org/10.1016/j.ehb.2019.01.006

Dare, S., Mackay, D. F., & Pell, J. P. (2015). Relationship between smoking and obesity: A cross-sectional study of 499,504 middle-aged adults in the UK general population. *PLoS ONE*, *10*(4), 1–12. https://doi.org/10.1371/journal.pone.0123579

Dartanto, T., Nurhasana, R., Thabrany, H., Moeis, F. R., & Satrya, A. (2018). *Parental Smoking Behavior and its Impact on Stunting , Cognitive , and Poverty: Empirical Evidence from the IFLS Panel Data*. Depok.

Davis, L., Barnes, A. J., Gross, A. C., Ryder, J. R., & Shlafer, R. J. (2019). Adverse Childhood Experiences and Weight Status among Adolescents. *Journal of Pediatrics*, *204*, 71-76.e1. https://doi.org/10.1016/j.jpeds.2018.08.071

Dinour, L., Leung, M. M., Tripicchio, G., Khan, S., & Yeh, M. C. (2012). The association between marital transitions, body mass index, and weight: A review of the literature. *Journal of Obesity*, *2012*, 1–16. https://doi.org/10.1155/2012/294974

Fang, H., Ali, M. M., & Rizzo, J. A. (2009). Does smoking affect body weight and obesity in China? *Economics and Human Biology*, *7*(3), 334–350.

Filozof, C., Fernández Pinilla, M. C., & Fernández-Cruz, A. (2004). Smoking cessation and weight gain. *Obesity Reviews*, *5*(2), 95–103. https://doi.org/10.1111/j.1467-789X.2004.00131.x

Flegal, K. M. (2007). The effects of changes in smoking prevalence on obesity prevalence in the United States. *American Journal of Public Health*, *97*(8), 1510–1514.

Ford, E. S., Anda, R. F., Edwards, V. J., Perry, G. S., Zhao, G., Li, C., & Croft, J. B. (2011). Adverse childhood experiences and smoking status in five states. *Preventive Medicine*, *53*(3), 188–193. https://doi.org/10.1016/j.ypmed.2011.06.015

Frankenberg, E., & Thomas, D. (2000). *The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2* (DRU-2238/1-NIA/NICHD).

Froom, P., Kristal-Boneh, E., Melamed, S., Gofer, D., Benbassat, J., & Ribak, J. (1999). Smoking cessation and body mass index of occupationally active men: The Israeli CORDIS study. *American Journal of Public Health*, *89*(5), 718–722.

https://doi.org/10.2105/AJPH.89.5.718

Grabowski, D. C., & Hirth, R. A. (2003). Competitive spillovers across non-profit and for-profit nursing homes. *Journal of Health Economics*, *22*, 1–22.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health, *80*(2), 223–255.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, *11*(1), 1–21.

Gruber, J., & Frakes, M. (2006). Does falling smoking lead to rising obesity? *Journal of Health Economics*, *25*, 183–197.

Guerra, F., Stringhini, S., Vollenweider, P., Waeber, G., & Marques-Vidal, P. (2015). Socio-demographic and behavioural determinants of weight gain in the Swiss population. *BMC Public Health*, *15*(1), 1–7.

Haj Mouhamed, D., Ezzaher, A., Hellara, I., Neffati, F., Douki, W., Gaha, L., & Najjar, M. F. (2013). Effect of cigarette smoking on insulin resistance risk. *European Psychiatry*, *28*, 1. https://doi.org/10.1016/s0924-9338(13)76917-7

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, *46*(6), 1251–1271.

Helble, M., & Francisco, K. (2017). *The imminent obesity crisis in Asia and the Pacific: First cost estimates* (No. 743). *ADBI Working Paper 743*. Tokyo. Retrieved from http://www.adb.org/

Hruby, A., Manson, J. E., Qi, L., Malik, V. S., Rimm, E. B., Sun, Q., ... Hu, F. B. (2016). Determinants and consequences of obesity. *American Journal of Public Health*, *106*(9), 1656–1662.

Huffman, S. K., & Rizov, M. (2007). Determinants of obesity in transition economies: The case of Russia. *Economics and Human Biology*, *5*, 379–391. https://doi.org/10.1016/j.ehb.2007.07.001

ILO. (2017). *Indonesia Jobs Outlook 2017: Harnessing technology for growth and job creation*. Jakarta.

Indonesia Investments. (2013). Indonesia's Most Populous Island Java Continues to Dominate the Economy. Retrieved July 30, 2019, from https://www.indonesia-investments.com/id/news/todays-headlines/indonesias-most-populous-island-of-java-continues-to-dominate-the-economy/item972

Indonesia National Research and Inovation Agency. (2016). Increase the publishment

of Research Result. Retrieved October 28, 2019, from https://international.ristekdikti.go.id/news/increase-the-publishment-of-research-result/

Indonesian Ministry of Health. (2018). *General Condition of Tobacco Consumption in Indonesia (Situasi Umum Konsumsi Tembakau di Indonesia). Infodatin (Pusat Data dan Informasi Kementerian Kesehatan RI).* Jakarta. Retrieved from file:///D:/infodatin tembakau per halaman.pdf

John, R. M. (2008). Crowding out effect of tobacco expenditure and its implications on household resource allocation in India. *Social Science and Medicine*, *66*(6), 1356–1367.

John, U., Hanke, M., Rumpf, H. J., & Thyrian, J. R. (2005). Smoking status, cigarettes per day, and their relationship to overweight and obesity among former and current smokers in a national adult general population sample. *International Journal of Obesity*, *29*(10), 1289–1294. https://doi.org/10.1038/sj.ijo.0803028

Jolliffe, D. (2011). Overweight and poor? on the relationship between income and the body mass index. *Economics and Human Biology*, *9*(4), 342–355. https://doi.org/10.1016/j.ehb.2011.07.004

Jumrani, J., & Birthal, P. S. (2017). Does consumption of tobacco and alcohol affect household food security? Evidence from rural India. *Food Security*, *9*(2), 255–279.

Jura, M., & Kozak, L. P. (2016). Obesity and related consequences to ageing. *Age*, *38*(23), 1–18.

Kamaura, M., Fujii, H., Mizushima, S., & Tochikubo, O. (2011). Weight gain and risk of impaired fasting glucose after smoking cessation. *Journal of Epidemiology*, *21*(6), 431–439. https://doi.org/10.2188/jea.JE20110010

Kasteridis, P., & Yen, S. T. (2012). Smoking cessation and body weight: Evidence from the behavioral risk factor surveillance survey. *Health Services Research*, *47*(4), 1580–1602. https://doi.org/10.1111/j.1475-6773.2012.01380.x

Kaufman, A., Augustson, E. M., & Patrick, H. (2012). Unraveling the relationship between smoking and weight: The role of sedentary behavior. *Journal of Obesity*, 1–10. https://doi.org/10.1155/2012/735465

Kim, D. (2018). Association Between Subjective Obesity Status and Smoking Behavior Among Normal-Weight Women. *Health Education & Behavior*, *45*(3), 394–400.

Køster-Rasmussen, R., Permin, C. A., Siersma, V., Henriksen, J. E., Heitmann, B. L.,

Heldgaard, P. E., & de Fine Olivarius, N. (2015). Back on track-Smoking cessation and weight changes over 9 years in a community-based cohort study. *Preventive Medicine*, *81*, 320–325. https://doi.org/10.1016/j.ypmed.2015.09.025

Kvaavik, E., Meyer, H. E., & Tverdal, A. (2004). Food habits, physical activity and body mass index in relation to smoking status in 40-42 year old Norwegian women and men. *Preventive Medicine*, *38*, 1–5. https://doi.org/10.1016/j.ypmed.2003.09.020

Lakdawalla, D., & Philipson, T. (2009). The growth of obesity and technological change. *Economics and Human Biology*, *7*(3), 283–293. https://doi.org/10.1016/j.ehb.2009.08.001

Lian, T. Y., & Dorotheo, U. (2018). *The Tobacco Control Atlas: ASEAN Region*. Bangkok.

Liang, L. L., & Mirelman, A. J. (2014). Why do some countries spend more for health? An assessment of sociopolitical determinants and international aid for government health expenditures. *Social Science and Medicine*, *114*, 161–168.

Liao, C., Gao, W., Cao, W., Lv, J., Yu, C., Wang, S., ... Li, L. (2018). Association of Educational Level and Marital Status with Obesity: A Study of Chinese Twins. *Twin Research and Human Genetics*, *21*(2), 126–135. https://doi.org/10.1017/thg.2018.8

Melchior, M., Chastang, J. F., Mackinnon, D., Galéra, C., & Fombonne, E. (2010). The intergenerational transmission of tobacco smoking-The role of parents' long-term smoking trajectories. *Drug and Alcohol Dependence*, *107*(2–3), 257–260.

Mendoza-Romero, D., Urbina, A., Cristancho-Montenegro, A., & Rombaldi, A. (2019). Impact of smoking and physical inactivity on self-rated health in women in Colombia. *Preventive Medicine Reports*, *16*(August), 100976.

Milner, A., Aitken, Z., Kavanagh, A., Lamontagne, A. D., Pega, F., & Petrie, D. (2017). Combining fixed effects and instrumental variable approaches for estimating the effect of psychosocial job quality on mental health: Evidence from 13 waves of a nationally representative cohort study. *Journal of Public Health*, *40*(2), 426–434.

Mizoue, T., Ueda, R., Tokui, N., Hino, Y., & Yoshimura, T. (1998). Body mass decrease after initial gain following smoking cessation. *International Journal of Epidemiology*, *27*(6), 984–988. https://doi.org/10.1093/ije/27.6.984

Morris, S. (2006). Body mass index and occupational attainment. *Journal of Health Economics*, *25*, 347–364.

Morris, S. (2007). The Impact of Obesity on Employment. *Labour Economics*, *14*(3), 413–433.

Munafò, M. R., Tilling, K., & Ben-Shlomo, Y. (2009). Smoking status and body mass index: A longitudinal study. *Nicotine and Tobacco Research*, *11*(6), 765–771. https://doi.org/10.1093/ntr/ntp062

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, *46*(1), 69–85.

Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., ... Gakidou, E. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *384*, 766–781. https://doi.org/10.1016/S0140-6736(14)60460-8

Nonnemaker, J., Finkelstein, E., Engelen, M., Hoerger, T., & Farrelly, M. (2009). Have efforts to reduce smoking really contributed to the obesity epidemic? *Economic Inquiry*, *47*, 366–376.

O'Connor, R. J. (2012). Non-cigarette tobacco products: What have we learned and where are we headed? *Tobacco Control*, *21*(2), 181–190.

Palaniappan, U., Starkey, L. J., O'Loughlin, J., & Gray-Donald, K. (2001). Fruit and Vegetable Consumption Is Lower and Saturated Fat Intake Is Higher among Canadians Reporting Smoking. *The Journal of Nutrition*, *131*(7), 1952–1958. https://doi.org/10.1093/jn/131.7.1952

Pan, J., Lei, X., & Liu, G. G. (2016). Health Insurance and Health Status: Exploring the Causal Effect from A Policy Intervention. *Health Economics*, *25*, 1389–1402. https://doi.org/10.1002/hec.3225

Perkins, K. A., Epstein, L. H., Stiller, R. L., Fernstrom, M. H., Sexton, J. E., Jacob, R. G., & Solberg, R. (1991). Acute effects of nicotine on hunger and caloric intake in smokers and nonsmokers. *Psychopharmacology*, *103*, 103–109. https://doi.org/10.1007/BF02244083

Pieroni, L., & Salmasi, L. (2016). The effect of smoking habit changes on body weight: Evidence from the UK. *Economics and Human Biology*, *20*, 1–13.

Piirtola, M., Jelenkovic, A., Latvala, A., Sund, R., Honda, C., Inui, F., ... Silventoinen,

K. (2018). Association of current and former smoking with body mass index: A study of smoking discordant twin pairs from 21 twin cohorts. *PLOS ONE*, *13*(7). https://doi.org/10.1371/journal.pone.0200140

Pisinger, C., Nielsen, H. Ø., Kuhlmann, C., & Rosthøj, S. (2017). Obesity Might Be a Predictor of Weight Reduction after Smoking Cessation. *Journal of Obesity*, *2017*. https://doi.org/10.1155/2017/2504078

Plurphanswat, N., & Rodu, B. (2014). The association of smoking and demographic characteristics on body mass index and obesity among adults in the U.S., 1999 – 2012. *BMC Obesity*, *1*(18), 1–9.

Pu, C. yun, Lan, V., Chou, Y. J., & Lan, C. fu. (2008). The crowding-out effects of tobacco and alcohol where expenditure shares are low: Analyzing expenditure data for Taiwan. *Social Science and Medicine*, *66*(9), 1979–1989.

Pulungan, A. B., Madarina, J., Batubara, J. R., & Hermanussen, M. (2018). Indonesian National Synthetic Growth Charts. *Acta Scientific Paediatrics*, *1*(1), 20–34.

Qi, Q., Chu, A. Y., Kang, J. H., Huang, J., Rose, L. M., Jensen, M. K., ... Qi, L. (2014). Fried food consumption, genetic risk, and body mass index: Gene-diet interaction analysis in three US cohort studies. *BMJ (Online)*, *348*(March), 1–12. https://doi.org/10.1136/bmj.g1610

Rachmi, C. N., Li, M., & Alison Baur, L. (2017). Overweight and obesity in Indonesia: prevalence and risk factors—a literature review. *Public Health*, *147*, 20–29. https://doi.org/10.1016/j.puhe.2017.02.002

Rashad, B. I., & Grossman, M. (2004). The Economics of Obesity. *Obesity*, *156*, 104–112.

Reas, D. l., Nygård, J. F., & Sørensen, T. (2009). Do quitters have anything to lose? Changes in body mass index for daily, never, and former smokers over an 11-year period (1990—2001). *Scandinavian Journal of Public Health*, *37*(7), 774–777. https://doi.org/10.1177/1403494809344654

Rehkopf, D. H., Headen, I., Hubbard, A., Deardorff, J., Kesavan, Y., Cohen, A. K., ... Abrams, B. (2016). Adverse childhood experiences and later life adult obesity and smoking in the United States. *Annals of Epidemiology*, *26*(7), 488–492.

Reid, H. H., & Ledgerwood, D. M. (2015). Depression, nicotine withdrawal, and smoking urges in smoking cessation. *Drug and Alcohol Dependence*, *146*(2015), e78–e79. https://doi.org/10.1016/j.drugalcdep.2014.09.582

Reitsma, M. B., Fullman, N., Ng, M., Salama, J. S., Abajobir, A., Abate, K. H., ... Gakidou, E. (2017). Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. *The Lancet*, *389*, 1885–1906.

Ritchie, H., & Roser, M. (2019). Obesity. Retrieved September 5, 2019, from https://ourworldindata.org/obesity

Robb, C. A., Huston, S. J., & Finke, M. S. (2008). The mitigating influence of time preference on the relation between smoking and BMI scores. *International Journal of Obesity*, *32*(11), 1670–1677. https://doi.org/10.1038/ijo.2008.151

Rodriguez, J., Jiang, R., Johnson, W. C., MacKenzie, B. A., Smith, L. J., & Barr, R. G. (2010). The Association of Pipe and Cigar Use With Cotinine Levels, Lung Function, and Airflow Obstruction. *Annals of Internal Medicine*, *152*, 201–210.

Sasso, A. T. Lo, & Buchmueller, T. C. (2004). The effect of the state children's health insurance program on health insurance coverage. *Journal of Health Economics*, *23*, 1059–1082. https://doi.org/10.1016/j.jhealeco.2004.03.006

Schaffer, M. E. ., & Stillman, S. (2010). Xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2 and xthtaylor. *Statistical Software Components*, (S456779). Retrieved from http://ideas.repec.org/c/boc/bocode/s456779.html

Scherr, A., Seifert, B., Kuster, M., Meyer, A., Fagerstroem, K. O., Tamm, M., & Stolz, D. (2015). Predictors of marked weight gain in a population of health care and industrial workers following smoking cessation. *BMC Public Health*, *15*(1), 1–11. https://doi.org/10.1186/s12889-015-1854-7

Schindler, T. H., Cardenas, J., Prior, J. O., Facta, A. D., Kreissl, M. C., Zhang, X. L., ... Schelbert, H. R. (2006). Relationship between increasing body weight, insulin resistance, inflammation, adipocytokine leptin, and coronary circulatory function. *Journal of the American College of Cardiology*, *47*(6), 1188–1195. https://doi.org/10.1016/j.jacc.2005.10.062

Sloan, F. A., Picone, G. A., Taylor Jr., D. H., & Chou, S.-Y. (2001). Hospital ownership and cost and quality of care: is there a dime's worth of difference? *Journal of Health Economics*, *20*, 1–21.

Sobal, J., & Hanson, K. L. (2011). Marital Status, Marital History, Body Weight, and Obesity. *Marriage and Family Review*, *47*(7), 474–504.

https://doi.org/10.1080/01494929.2011.620934

Sohn, K. (2015). The effects of smoking on obesity: evidence from Indonesian panel data. *Tobacco Induced Diseases*, *13*(39), 1–8.

Statistics Indonesia. (2017). *Concept and Definition of National Socio-economic Survey 2017. Statistics Indonesia*.

Stock, J. H., & Yogo, M. (2005). Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments. In D. W. K. Andrews & J. H. E. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 109–120). Cambridge University Press. https://doi.org/10.1017/CBO9780511614491.007

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (2004). *The Third Wave of the Indonesia Family Life Survey: Overview and Field Report* (WR144/1-NIA/NICHD) (Vol. 1).

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herwati, Y., Witoelar, F., & Corporation, R. (2016). *The 5th Wave of the Indonesia Family Life Survey (IFLS): Overview and Field Report*. Retrieved from http://www.rand.org/content/dam/rand/pubs/working_papers/WR1100/WR1143z2/RAND_WR1143z2.pdf

Strauss, J., Witoelar, F., & Sikoki, B. (2016). *The Fifth Wave of the Indonesia Family Life Survey: Overview and Field Report* (Vol. 1). Santa Monica.

Strauss, J., Witoelar, F., Sikoki, B., & Wattie, A. M. (2009). *The Fourth Wave of the Indonesia Family Life Survey: Overview and Field Report* (WR-675/1-NIA/NICHD) (Vol. 1).

Tafreschi, D. (2015). The income body weight gradients in the developing economy of China. *Economics and Human Biology*, *16*, 115–134. https://doi.org/10.1016/j.ehb.2014.02.001

Taleb, Z. Ben, Ward, K. D., Asfar, T., Jaber, R., Bahelah, R., & Maziak, W. (2017). Smoking cessation and changes in body mass index: Findings from the first randomized cessation trial in a low-income country setting. *Nicotine and Tobacco Research*, *19*(3), 351–356. https://doi.org/10.1093/ntr/ntw223

Tandon, A., Pambudi, E., Harimurti, P., Masaki, E., Subandoro, A., Marzoeki, P., ... Suharno, N. (2016). *Indonesia Health Financing System Assessment: Spend More, Right & Better*.

Tanumihardjo, S. A., Anderson, C., Kaufer-Horwitz, M., Bode, L., Emenaker, N. J., Haqq, A. M., ... Stadler, D. D. (2007). Poverty, Obesity, and Malnutrition: An International Perspective Recognizing the Paradox. *Journal of the American Dietetic Association*, *107*(11), 1966–1972. https://doi.org/10.1016/j.jada.2007.08.007

Tekin, E., & Markowitz, S. (2008). The Relationship between Suicidal Behavior and Productive Activities of Young Adults. *Southern Economic Journal*, *75*(2), 300–331.

Tremmel, M., Gerdtham, U. G., Nilsson, P. M., & Saha, S. (2017). Economic burden of obesity: A systematic literature review. *International Journal of Environmental Research and Public Health*, *14*(4), 1–18. https://doi.org/10.3390/ijerph14040435

Tzotzas, T., Vlahavas, G., Papadopoulou, S. K., Kapantais, E., Kaklamanou, D., & Hassapidou, M. (2010). Marital status and educational level associated to obesity in Greek adults: Data from the National Epidemiological Survey. *BMC Public Health*, *10*(732), 1–8. https://doi.org/10.1186/1471-2458-10-732

Veldheer, S., Yingst, J., Zhu, J., & Foulds, J. (2015). Ten-year weight gain in smokers who quit, smokers who continued smoking and never smokers in the United States, NHANES 2003-2012. *International Journal of Obesity*, *39*(12), 1727–1732. https://doi.org/10.1038/ijo.2015.127

Wang, Q. (2015). Smoking and body weight: Evidence from China health and nutrition survey. *BMC Public Health*, *15*(1), 1–8.

Watanabe, T., Tsujino, I., Konno, S., Ito, Y. M., Takashina, C., Sato, T., ... Nishimura, M. (2016). Association between smoking status and obesity in a nationwide survey of Japanese adults. *PLoS ONE*, *11*(3), 1–14. https://doi.org/10.1371/journal.pone.0148926

Webb, P., & Bain, C. (2011). *Essential Epidemiology: An Introduction for Students and Health Professionals* (2nd ed.). Cambridge: Cambridge University Press.

Wehby, G. L., & Courtemanche, C. J. (2012). The heterogeneity of the cigarette price effect on body mass index. *Journal of Health Economics*, *31*(5), 719–729.

Wehby, G. L., Murray, J. C., Wilcox, A., & Lie, R. T. (2012). Smoking and body weight: Evidence using genetic instruments. *Economics and Human Biology*, *10*, 113–126.

Weinberger, A. H., Kashan, R. S., Shpigel, D. M., Esan, H., Taha, F., Lee, C. J., ...

Goodwin, R. D. (2017). Depression and cigarette smoking behavior: A critical review of population-based studies. *The American Journal of Drug and Alcohol Abuse*, *43*(4), 416–431. https://doi.org/10.3109/00952990.2016.1171327

WHO. (2008). Tobacco Free Initiative. Retrieved October 22, 2019, from http://www.emro.who.int/tfi/mpower/index.html

WHO. (2012). *Global Adult Tobacco Survey: Indonesia Report 2011*. Jakarta: World Health Organization.

WHO. (2014). *Global Status Report on Noncommunicable Diseases 2014. World Health Organization*.

WHO. (2016). Diabetes country profiles 2016: Indonesia. Retrieved November 16, 2018, from http://www.who.int/diabetes/country-profiles/idn_en.pdf

WHO expert consultation. (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*, *363*, 157–163. https://doi.org/10.1016/S0140-6736(03)15268-3

WHO Regional Office for South-East Asia. (2018). *Indonesia Tobacco Factsheet 2018*.

Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach* (Sixth Edit). Boston: Cengage Learning.

World Bank. (2019). Smoking Prevalence. Retrieved May 6, 2019, from https://data.worldbank.org/indicator/SH.PRV.SMOK.FE

World Bank Group. (2018). *The Economics of Tobacco Taxation and Employment in Indonesia*. Washington D.C.: The World Bank Group.

Xu, F., Yin, X.-M., & Wang, Y. (2007). The association between amount of cigarettes smoked and overweight, central obesity among Chinese adults in Nanjing, China. *Asia Pacific Journal of Clinical Nutrition*, *16*(2), 240–247.

Appendices

Appendix A. Hausman test of random effects versus fixed effects

A.1. Hausman test result for a structural equation with current smoking status

```
Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re  robust cluster(pidlink_num)
Sargan-Hansen statistic 2198.673  Chi-sq(11)  P-value = 0.0000
```

A.2. Hausman test result for a structural equation with the number of daily cigarettes smoked by current smokers

```
Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re  robust cluster(pidlink_num)
Sargan-Hansen statistic 2264.373  Chi-sq(11)  P-value = 0.0000
```

A.3. Hausman test result for a structural equation with duration of smoking of current smokers

```
Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re  robust cluster(pidlink_num)
Sargan-Hansen statistic 2194.390  Chi-sq(11)  P-value = 0.0000
```

A.4. Hausman test result for a structural equation with former smoking status

```
Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re  robust cluster(pidlink_num)
Sargan-Hansen statistic 1440.379  Chi-sq(11)  P-value = 0.0000
```

A.5. Hausman test result for a first-stage regression with daily cigarettes smoked by current smokers as the dependent variable

```
Test of overidentifying restrictions: fixed vs random effects
Cross-section time-series model: xtreg re  robust cluster(pidlink_num)
Sargan-Hansen statistic 2551.734  Chi-sq(9)   P-value = 0.0000
```

Appendix B. Endogeneity tests

B.1 Endogeneity test of current smoking status

```
Random-effects GLS regression                   Number of obs     =     69,220
Group variable: pidlink_num                     Number of groups  =     37,399

R-sq:                                           Obs per group:
     within  = 0.3201                                         min =          1
     between = 0.1505                                         avg =        1.9
     overall = 0.1738                                         max =          4

                                                Wald chi2(12)     =   15012.62
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000
```

(Std. Err. adjusted for 37,399 clusters in pidlink_num)

bmiclean_hw_gr	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
smoke_curr	-1.707943	.0412887	-41.37	0.000	-1.788868	-1.627019
resid	1.314723	.0957353	13.73	0.000	1.127085	1.502361
age	.355641	.005367	66.26	0.000	.345122	.3661601
age2	-.0034792	.0000621	-56.05	0.000	-.0036009	-.0033576
married	.4060294	.0357177	11.37	0.000	.3360239	.4760349
urban	.3166047	.0310892	10.18	0.000	.255671	.3775384
java	-.0425535	.0366214	-1.16	0.245	-.1143301	.029223
log_pcincome	.3763201	.0091496	41.13	0.000	.3583872	.3942531
log_pcsugar	-.0142454	.0141108	-1.01	0.313	-.0419021	.0134112
log_pcoil	.1541212	.0139617	11.04	0.000	.1267567	.1814857
total_weeklyhours	.0003281	.0004254	0.77	0.441	-.0005057	.0011618
year_schooling	.084128	.0046361	18.15	0.000	.0750415	.0932146
_cons	8.189645	.1796921	45.58	0.000	7.837455	8.541835
sigma_u	3.1546843					
sigma_e	1.8486755					
rho	.74437588	(fraction of variance due to u_i)				

B.2 Endogeneity test of current daily cigarettes smoked

```
Random-effects GLS regression                   Number of obs     =     69,149
Group variable: pidlink_num                     Number of groups  =     37,374

R-sq:                                           Obs per group:
     within  = 0.3186                                         min =          1
     between = 0.1365                                         avg =        1.9
     overall = 0.1596                                         max =          4

                                                Wald chi2(12)     =   14465.35
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000
```

(Std. Err. adjusted for 37,374 clusters in pidlink_num)

bmiclean_hw_gr	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
cig_current	-.1065378	.0036208	-29.42	0.000	-.1136344	-.0994412
resid_cig	.0994196	.0046159	21.54	0.000	.0903726	.1084665
age	.3631903	.0054152	67.07	0.000	.3525767	.3738039
age2	-.0035756	.0000626	-57.10	0.000	-.0036984	-.0034529
married	.4137285	.0358034	11.56	0.000	.3435552	.4839018
urban	.3367972	.0312617	10.77	0.000	.2755254	.3980689
java	-.1222122	.0370332	-3.30	0.001	-.194796	-.0496284
log_pcincome	.3939595	.0091916	42.86	0.000	.3759443	.4119747
log_pcsugar	-.0091229	.0141642	-0.64	0.520	-.0368841	.0186384
log_pcoil	.1597973	.0140092	11.41	0.000	.1323399	.1872548
total_weeklyhours	.0006361	.000438	1.45	0.146	-.0002222	.0014945
year_schooling	.0857206	.0046642	18.38	0.000	.076579	.0948621
_cons	7.577579	.1803119	42.02	0.000	7.224174	7.930984
sigma_u	3.1834849					
sigma_e	1.8489289					
rho	.74776733	(fraction of variance due to u_i)				

B.3 Endogeneity test of current duration of smoking

```
Random-effects GLS regression                   Number of obs     =     68,968
Group variable: pidlink_num                     Number of groups  =     37,329

R-sq:                                           Obs per group:
     within  = 0.3256                                         min =          1
     between = 0.1466                                         avg =        1.8
     overall = 0.1709                                         max =          4

                                                Wald chi2(12)     =   15044.87
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

                                (Std. Err. adjusted for 37,329 clusters in pidlink_num)
```

bmiclean_hw_gr	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
duration_curr	-.0696659	.0018746	-37.16	0.000	-.07334	-.0659918
resid_dur	.0334732	.0030142	11.11	0.000	.0275655	.0393809
age	.3484976	.0054184	64.32	0.000	.3378776	.3591175
age2	-.0031741	.0000632	-50.21	0.000	-.003298	-.0030501
married	.5214955	.036032	14.47	0.000	.4508741	.5921168
urban	.3118063	.031109	10.02	0.000	.2508338	.3727789
java	-.028323	.0367496	-0.77	0.441	-.100351	.043705
log_pcincome	.3793611	.0091796	41.33	0.000	.3613694	.3973529
log_pcsugar	-.0144743	.0141037	-1.03	0.305	-.042117	.0131684
log_pcoil	.1537188	.0139527	11.02	0.000	.126372	.1810656
total_weeklyhours	-.0009844	.0004214	-2.34	0.019	-.0018103	-.0001586
year_schooling	.0831365	.0046429	17.91	0.000	.0740366	.0922364
_cons	7.802979	.1796857	43.43	0.000	7.450802	8.155157
sigma_u	3.1631994					
sigma_e	1.8410776					
rho	.74696035	(fraction of variance due to u_i)				

Appendix C. Detection of weak instruments

C.1 Detection of weak instruments for current smoking status (full sample; all sexes)

```
Random-effects probit regression              Number of obs     =     89,507
Group variable: pidlink_num                   Number of groups  =     42,848

Random effects u_i ~ Gaussian                 Obs per group:
                                                              min =          1
                                                              avg =        2.1
                                                              max =          4

Integration method: mvaghermite               Integration pts.  =         12

                                              Wald chi2(18)     =    1331.12
Log pseudolikelihood  = -35749.817            Prob > chi2       =     0.0000
```

(Std. Err. adjusted for 42,848 clusters in pidlink_num)

smoke_curr	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
pcsmokers	.0411377	.0064419	6.39	0.000	.0285118	.0537636
mun_pcsmokers	.0711016	.0166889	4.26	0.000	.0383918	.1038113
age	.1719405	.0095301	18.04	0.000	.153262	.1906191
age2	-.0020897	.0001182	-17.67	0.000	-.0023214	-.0018579
married	-.1215588	.0507935	-2.39	0.017	-.2211122	-.0220054
urban	-.0191727	.0525816	-0.36	0.715	-.1222307	.0838853
java	-.35478	.1851219	-1.92	0.055	-.7176122	.0080522
log_pcincome	.0365805	.0148939	2.46	0.014	.0073891	.0657719
total_weeklyhours	.0058072	.0005349	10.86	0.000	.0047589	.0068555
year_schooling	.0975589	.0107228	9.10	0.000	.0765426	.1185751
mun_age	-.2051961	.0166758	-12.31	0.000	-.2378801	-.1725121
mun_age2	.0026499	.000198	13.39	0.000	.0022619	.0030379
mun_married	-.0741441	.1009658	-0.73	0.463	-.2720334	.1237453
mun_urban	-.4059805	.093775	-4.33	0.000	-.5897761	-.2221849
mun_java	.2916621	.1962517	1.49	0.137	-.0929842	.6763084
mun_logpcincome	-.2108912	.0331792	-6.36	0.000	-.2759212	-.1458612
mun_hours	.0809761	.0057589	14.06	0.000	.0696888	.0922634
mun_yearschooling	-.0681771	.0149818	-4.55	0.000	-.0975409	-.0388133
_cons	-6.07032	.6540782	-9.28	0.000	-7.352289	-4.78835
/lnsig2u	2.99876	.			.	.
sigma_u	4.478912	.			.	.
rho	.9525181	.			.	.

C.2 Detection of weak instruments for the number of daily cigarettes smoked by current smokers (CSNS sample; all sexes)

```
Fixed-effects (within) regression             Number of obs     =     84,533
Group variable: pidlink_num                   Number of groups  =     42,042

R-sq:                                         Obs per group:
     within  = 0.0298                                         min =          1
     between = 0.0403                                         avg =        2.0
     overall = 0.0309                                         max =          4

                                              F(9,42041)        =     102.02
corr(u_i, Xb)  = 0.0593                       Prob > F          =     0.0000
```

(Std. Err. adjusted for 42,042 clusters in pidlink_num)

clean_cig	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
mean_cig_m1	.1067694	.0228191	4.68	0.000	.0620434	.1514954
age	.1800654	.0097732	18.42	0.000	.1609097	.1992212
age2	-.0021044	.0001099	-19.14	0.000	-.0023199	-.0018889
married	.0337219	.0590841	0.57	0.568	-.0820842	.149528
urban	.0421846	.0673849	0.63	0.531	-.0898913	.1742605
java	-.5277208	.2686187	-1.96	0.049	-1.054219	-.0012227
log_pcincome	.126543	.0181033	6.99	0.000	.0910601	.1620258
total_weeklyhours	.0060835	.0007024	8.66	0.000	.0047067	.0074603
year_schooling	.0355423	.0135798	2.62	0.009	.0089256	.062159
_cons	-2.549026	.3495808	-7.29	0.000	-3.234212	-1.863841
sigma_u	6.2140325					
sigma_e	3.1876812					
rho	.79167186	(fraction of variance due to u_i)				

C.3 Detection of weak instruments for duration of smoking of current smokers (CSNS sample; all sexes)

```
Fixed-effects (within) regression               Number of obs      =     84,555
Group variable: pidlink_num                     Number of groups   =     42,054

R-sq:                                           Obs per group:
     within  = 0.1666                                         min =          1
     between = 0.1784                                         avg =        2.0
     overall = 0.1505                                         max =          4

                                                F(10,42053)        =     534.96
corr(u_i, Xb)  = 0.0218                         Prob > F           =     0.0000

                              (Std. Err. adjusted for 42,054 clusters in pidlink_num)
```

clean_dc	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
mean_dc1	.0252128	.0531212	0.47	0.635	-.0789058	.1293314
smoke_expodurr1	-.001132	.0016485	-0.69	0.492	-.0043631	.0020992
age	.2315166	.0407413	5.68	0.000	.1516628	.3113704
age2	.0009868	.0002372	4.16	0.000	.0005219	.0014516
married	.7436856	.0848734	8.76	0.000	.577332	.9100391
urban	-.06719	.0976512	-0.69	0.491	-.2585883	.1242083
java	-.6281657	.2887758	-2.18	0.030	-1.194172	-.0621593
log_pcincome	.0207832	.0289753	0.72	0.473	-.036009	.0775754
total_weeklyhours	-.0041671	.0009725	-4.29	0.000	-.0060732	-.0022611
year_schooling	-.0853159	.0200096	-4.26	0.000	-.124535	-.0460967
_cons	-2.457009	1.195382	-2.06	0.040	-4.799982	-.1140359
sigma_u	10.631589					
sigma_e	4.6342296					
rho	.84033455	(fraction of variance due to u_i)				

```
.
end of do-file

. test mean_dc1 smoke_expodurr1

 ( 1)  mean_dc1 = 0
 ( 2)  smoke_expodurr1 = 0

       F(  2, 42053) =    0.38
            Prob > F =    0.6836

.
```

Appendix D. First-stage regressions

D.1 First-stage regression of current smoking status: full sample

Average Marginal Effects of Probit with Random Effects: Full Sample

	All	Male	Female
Percentage of current smokers	0.00551***	0.00790***	0.00108***
	(0.000487)	(0.000761)	(0.000245)
Age	0.0137***	0.0251***	0.00323***
	(0.000526)	(0.000820)	(0.000298)
Age Square	-0.000145***	-0.000284***	-0.0000262***
	(0.00000612)	(0.00000914)	(0.00000300)
Married	-0.0107***	-0.00122	-0.00414**
	(0.00351)	(0.00619)	(0.00163)
Urban	-0.0258***	-0.0356***	0.00594***
	(0.00332)	(0.00512)	(0.00167)
Java	-0.00812**	-0.00139	-0.00351**
	(0.00387)	(0.00577)	(0.00163)
Log effective per capita income	-0.00560***	-0.00542***	-0.000973**
	(0.000973)	(0.00154)	(0.000488)
Hours of work	0.00190***	0.00159***	0.0000441**
	(0.0000393)	(0.0000750)	(0.0000222)
Years of schooling	0.00177***	-0.0121***	-0.000718***
	(0.000460)	(0.000716)	(0.000211)
Observations	89507	41730	46632

Cluster-robust Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D.2 First-stage Regression of Current Smoking Status: Current-smoker and Non-smoker (CSNS)

Average marginal effects of probit with random effects: CSNS sample

	All	Male	Female
Percentage of current smokers	0.00553***	0.00377***	0.00105***
	(0.000643)	(0.000728)	(0.000247)
Age	0.0198***	0.0156***	0.00343***
	(0.00114)	(0.00176)	(0.000377)
Age square	-0.000190***	-0.000159***	-0.0000272***
	(0.0000123)	(0.0000178)	(0.00000382)
Married	0.0333***	0.00540	-0.00314
	(0.00542)	(0.00555)	(0.00191)
Urban	-0.0181***	-0.0196***	0.00676***
	(0.00506)	(0.00498)	(0.00179)
Java	0.0107	-0.00347	-0.00387*
	(0.00719)	(0.00553)	(0.00209)
Log effective per capita income	-0.00972***	-0.00434**	-0.00142***
	(0.00141)	(0.00200)	(0.000481)
Hours of work	0.00180***	0.000767***	0.0000535**
	(0.0000638)	(0.000102)	(0.0000236)
Years of schooling	0.00433***	-0.00611***	-0.000849***
	(0.000749)	(0.00142)	(0.000263)
Observations	84904	37761	46099
chi2			

Cluster-robust Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D.3 First-stage regression of current smoking status: current-smoker and former-smoker (CSFS) sample

Average marginal effects of probit with random effects: CSFS sample

	All	Male	Female
Percentage of current smokers	0.00502***	0.00470***	0.0140***
	(0.000698)	(0.000699)	(0.00464)
Age	0.00820***	0.00887***	0.0288***
	(0.000715)	(0.000729)	(0.00435)
Age square	-0.000116***	-0.000124***	-0.000276***
	(0.00000767)	(0.00000779)	(0.0000448)
Married	-0.00746	-0.00784	-0.0586**
	(0.00553)	(0.00578)	(0.0289)
Urban	-0.0228***	-0.0187***	-0.00247
	(0.00444)	(0.00443)	(0.0319)
Java	0.00531	0.00679	-0.0183
	(0.00429)	(0.00431)	(0.0260)
Log effective per capita income	-0.00130	-0.00142	-0.000749
	(0.00146)	(0.00146)	(0.00951)
Hours of work	0.00134***	0.00121***	0.000520
	(0.0000732)	(0.0000746)	(0.000427)
Years of schooling	-0.00362***	-0.00497***	0.00504
	(0.000569)	(0.000578)	(0.00366)
Observations	33497	31092	1542

Cluster-robust standard errors in parentheses
* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$

D.4 First-stage regression of the number of daily cigarettes smoked by current smokers: current-smoker and non-smoker (CSNS) sample

First-stage regression of the number of daily cigarettes smoked by current smokers: CSNS sample

	All	Male	Female
Community's average daily cigarettes smoked by current smokers	0.106***	0.230***	0.00949
	(0.0228)	(0.0521)	(0.00644)
Age	0.179***	0.446***	0.00689***
	(0.00975)	(0.0241)	(0.00257)
Age square	-0.00209***	-0.00517***	-0.0000720**
	(0.000110)	(0.000263)	(0.0000318)
Married	0.0359	-0.125	-0.00980
	(0.0590)	(0.154)	(0.0154)
Urban	0.0473	-0.000721	0.0300*
	(0.0672)	(0.155)	(0.0168)
Java	-0.529**	-0.443	-0.00963
	(0.269)	(0.566)	(0.0371)
Log effective per capita income	0.125***	0.325***	-0.00368
	(0.0181)	(0.0440)	(0.00456)
Hours of work	0.00613***	0.0122***	0.000515**
	(0.000701)	(0.00167)	(0.000206)
Years of schooling	0.0359***	0.0742**	0.0000679
	(0.0136)	(0.0290)	(0.00514)
Constant	-2.527***	-7.682***	-0.0595
	(0.350)	(0.804)	(0.0869)
Observations	84487	37377	46079

Cluster-robust Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D.5 First-stage regression of the number of daily cigarettes smoked by current smokers: current-smoker and former-smoker (CSFS) sample

First-stage regression of the number of daily cigarettes smoked by current smokers: CSFS sample

	All	Male	Female
Community's average daily cigarettes smoked by current smokers	0.153**	0.162**	-0.0191
	(0.0627)	(0.0654)	(0.233)
Age	0.589***	0.605***	0.471***
	(0.0283)	(0.0295)	(0.112)
Age square	-0.00707***	-0.00734***	-0.00467***
	(0.000311)	(0.000327)	(0.00105)
Married	-0.407**	-0.448**	0.256
	(0.184)	(0.193)	(0.677)
Urban	-0.0120	-0.0395	1.032
	(0.190)	(0.198)	(0.681)
Java	-0.611	-0.516	-6.425***
	(0.682)	(0.705)	(1.690)
Log effective per capita income	0.349***	0.369***	0.171
	(0.0521)	(0.0547)	(0.187)
Hours of work	0.0207***	0.0213***	0.00185
	(0.00194)	(0.00202)	(0.00717)
Years of schooling	0.238***	0.249***	-0.0286
	(0.0373)	(0.0387)	(0.157)
Constant	-9.152***	-9.641***	-5.126
	(0.983)	(1.024)	(3.724)
Observations	33058	30688	1520

Cluster-robust Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D.6 First-stage regression of the number of daily cigarettes smoked by current smokers: current-smoker (CS) sample

First-stage regression of the number of daily cigarettes smoked by current smokers: current-smoker sample

	All	Male	Female
Community's average daily cigarettes smoked by current smokers	0.339***	0.346***	0.513*
	(0.0654)	(0.0673)	(0.309)
Age	0.442***	0.451***	0.138
	(0.0297)	(0.0307)	(0.145)
Age square	-0.00520***	-0.00534***	-0.00160
	(0.000321)	(0.000333)	(0.00115)
Married	-0.218	-0.243	0.395
	(0.190)	(0.197)	(0.814)
Urban	0.0711	0.0515	0.413
	(0.198)	(0.205)	(1.071)
Java	-0.101	-0.166	0
	(0.717)	(0.736)	(.)
Log effective per capita income	0.392***	0.402***	0.146
	(0.0526)	(0.0544)	(0.248)
Hours of work	0.0127***	0.0129***	0.00850
	(0.00202)	(0.00209)	(0.0100)
Years of schooling	0.119***	0.127***	-0.172
	(0.0396)	(0.0407)	(0.196)
Constant	-7.044***	-7.278***	-2.783
	(1.035)	(1.063)	(5.132)
Observations	28480	26741	990

Cluster-robust Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix E. Second-stage regressions with fixed effects and instrumental variables (FEIV) of current smoking status

E.1 Second-stage regression of current smoking status on BMI: full sample

Second-stage regression of current smoking status on BMI: full sample

	All (1)	All (2)	Male (1)	Male (2)	Female (1)	Female (2)
Current smoking status	-7.060	-4.321	3.480	4.700	-81.52	-64.03
	(4.615)	(4.274)	(2.961)	(3.145)	(75.50)	(55.33)
Former smoking status	-4.879	-2.901	2.750	3.636	-51.26	-40.25
	(3.338)	(3.090)	(2.151)	(2.285)	(47.59)	(34.88)
Age	0.480***	0.457***	0.275***	0.254***	0.581***	0.560***
	(0.0369)	(0.0344)	(0.0492)	(0.0524)	(0.0937)	(0.0707)
Age square	-0.00400***	-0.00376***	-0.00206***	-0.00184***	-0.00470***	-0.00452***
	(0.000400)	(0.000371)	(0.000528)	(0.000562)	(0.000933)	(0.000709)
Married	0.158***	0.150***	0.0819	0.0791	0.181	0.208
	(0.0531)	(0.0493)	(0.0691)	(0.0730)	(0.254)	(0.201)
Hours of work	0.00324**	0.00236	-0.000555	-0.00118	0.00408	0.00330
	(0.00156)	(0.00144)	(0.00179)	(0.00190)	(0.00366)	(0.00277)
Years of schooling	0.0574**	0.0427*	0.0475	0.0344	-0.0540	-0.0510
	(0.0238)	(0.0220)	(0.0313)	(0.0332)	(0.0508)	(0.0404)
Log effective per capita income	0.0914***	0.0859***	0.0675***	0.0618***	0.144*	0.128**
	(0.0151)	(0.0140)	(0.0185)	(0.0197)	(0.0844)	(0.0653)
Urban	0.00927	-0.00651	0.0648	0.0524	0.313	0.225
	(0.0500)	(0.0466)	(0.0654)	(0.0695)	(0.378)	(0.285)
Java	-0.491**	-0.378*	0.0224	0.119	-1.375	-1.174

	(0.233)	(0.216)	(0.281)	(0.303)	(1.234)	(0.959)
Log per capita oil cons.		0.0775***		0.0681***		0.0958
		(0.0169)		(0.0233)		(0.0669)
Log per capita sugar cons.		0.0475***		0.0554**		0.0865
		(0.0168)		(0.0239)		(0.0915)
Constant	12.03***	10.53***	10.86***	9.424***	10.55***	8.966***
	(0.706)	(0.669)	(0.830)	(0.917)	(0.992)	(1.085)
Observations	72918	72918	33578	33578	38440	38440
chi2	352306.0	408300.2	222218.6	202326.6	16795.3	25726.5

Cluster-robust Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Endogenous variable: current smoking status
Instrumental variable of current smoking status: percentage of current smokers in the community

E.2 Second-stage regression of current smoking status on BMI: Current-smoker and non-smoker (CSNS) sample

Second-stage regression of current smoking status on BMI: CSNS sample

	All (1)	All (2)	Male (1)	Male (2)	Female (1)	Female (2)
Current smoking status	-9.094	-5.945	4.256	5.537	-117.3	-93.04
	(5.789)	(5.291)	(3.300)	(3.498)	(128.9)	(94.51)
Age	0.485***	0.466***	0.277***	0.258***	0.560***	0.546***
	(0.0332)	(0.0306)	(0.0464)	(0.0493)	(0.0947)	(0.0728)
Age square	-0.00405***	-0.00386***	-0.00211***	-0.00192***	-0.00446***	-0.00434***
	(0.000361)	(0.000331)	(0.000503)	(0.000533)	(0.000941)	(0.000731)
Married	0.152***	0.143***	0.0493	0.0444	0.258	0.265
	(0.0582)	(0.0534)	(0.0773)	(0.0823)	(0.295)	(0.237)
Hours of work	0.00284**	0.00214*	-0.000776	-0.00124	0.00598	0.00482
	(0.00137)	(0.00126)	(0.00153)	(0.00163)	(0.00606)	(0.00449)
Years of schooling	0.0603**	0.0444*	0.0328	0.0181	-0.0338	-0.0352
	(0.0279)	(0.0255)	(0.0366)	(0.0388)	(0.0545)	(0.0442)
Log effective per capita income	0.0894***	0.0842***	0.0499**	0.0429*	0.0896	0.0863
	(0.0162)	(0.0149)	(0.0211)	(0.0224)	(0.0730)	(0.0591)
Urban	0.0120	-0.00643	0.0288	0.0155	0.387	0.289
	(0.0552)	(0.0509)	(0.0720)	(0.0770)	(0.522)	(0.394)
Java	-0.578**	-0.451*	0.240	0.337	-1.937	-1.643
	(0.267)	(0.247)	(0.335)	(0.365)	(1.578)	(1.194)
Log per capita oil cons.		0.0774***		0.0802***		0.0799
		(0.0191)		(0.0258)		(0.0873)

Log per capita sugar cons.		0.0432**		0.0368		0.102
		(0.0192)		(0.0267)		(0.123)
Constant	12.61***	10.98***	10.60***	9.165***	11.87***	10.01***
	(1.073)	(0.984)	(1.082)	(1.176)	(1.722)	(1.127)
Observations	69220	69220	30388	30388	38010	38010
chi2	318892.2	367935.2	196654.5	180492.2	11226.6	17174.7

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Endogenous variable: current smoking status
Instrumental variable of current smoking status: percentage of current smokers in the community

E.3 Second-stage regression of current smoking status on BMI: current-smoker and former-smoker (CSFS) sample

Second-stage regression of current smoking status on BMI: CSFS sample

	All (1)	All (2)	Male (1)	Male (2)	Female (1)	Female (2)
Current smokers	0.188	0.708	0.157	0.657	-0.0149	0.997
	(1.138)	(1.150)	(1.241)	(1.260)	(3.091)	(2.765)
Age	0.292***	0.279***	0.293***	0.281***	0.365*	0.306
	(0.0313)	(0.0314)	(0.0331)	(0.0333)	(0.210)	(0.195)
Age square	-0.00227***	-0.00210***	-0.00230***	-0.00214***	-0.00275	-0.00217
	(0.000388)	(0.000391)	(0.000420)	(0.000424)	(0.00198)	(0.00181)
Married	0.0985	0.124	0.102	0.124	-0.168	-0.0935
	(0.0829)	(0.0835)	(0.0840)	(0.0846)	(0.344)	(0.338)
Hours of work	0.000729	0.000240	0.000476	-0.0000303	0.00372	0.00381
	(0.00134)	(0.00136)	(0.00149)	(0.00151)	(0.00403)	(0.00409)
Years of schooling	0.0419**	0.0337	0.0460**	0.0378	0.0641	0.0612
	(0.0206)	(0.0207)	(0.0231)	(0.0233)	(0.0735)	(0.0747)
Log effective per capita income	0.0622***	0.0569***	0.0638***	0.0588***	-0.0736	-0.0904
	(0.0176)	(0.0177)	(0.0186)	(0.0187)	(0.0942)	(0.0939)
Urban	0.0432	0.0369	0.0547	0.0503	-0.289	-0.359
	(0.0725)	(0.0735)	(0.0752)	(0.0763)	(0.406)	(0.429)
Java	0.105	0.209	0.137	0.236	0	0
	(0.302)	(0.298)	(0.308)	(0.304)	(.)	(.)
Log per capita oil cons.		0.0448**		0.0379*		0.165
		(0.0217)		(0.0228)		(0.116)

Log per capita sugar cons.		0.0733***		0.0717***		0.0648
		(0.0227)		(0.0241)		(0.141)
Constant	12.69***	11.47***	12.63***	11.48***	12.94***	11.76***
	(0.405)	(0.462)	(0.439)	(0.501)	(3.772)	(4.544)
Observations	27083	27083	25142	25142	1260	1260
chi2	183993.0	181711.1	182376.0	179923.2	3974.8	4222.3

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Endogenous variable: current smoking status
Instrumental variable of current smoking status: percentage of current smokers in the community

Appendix F. Second-stage regressions with fixed effects and instrumental variables (FEIV) of current smoking status

F.1 Second-stage regression of the number of daily cigarettes smoked by current smokers on BMI: current-smoker and non-smoker (CSNS) sample

Second-stage regression of the number of daily cigarettes smoked by current smokers on BMI: CSNS sample

	All (1)	All (2)	Male (1)	Male (2)
The number of daily cigarettes smoked by current smokers	0.725***	0.690***	-0.101	-0.155
	(0.222)	(0.235)	(0.0851)	(0.0980)
Age	0.302***	0.307***	0.381***	0.406***
	(0.0419)	(0.0443)	(0.0411)	(0.0471)
Age square	-0.00197***	-0.00204***	-0.00329***	-0.00357***
	(0.000485)	(0.000513)	(0.000467)	(0.000537)
Married	0.0881	0.0908	0.0663	0.0649
	(0.0683)	(0.0667)	(0.0726)	(0.0763)
Hours of work	-0.00363**	-0.00341**	0.00211	0.00285**
	(0.00155)	(0.00160)	(0.00129)	(0.00146)
Years of schooling	-0.00757	-0.00700	0.0875***	0.0899***
	(0.0165)	(0.0162)	(0.0139)	(0.0149)
Log effective per capita income	-0.00401	-0.000734	0.0978***	0.111***
	(0.0319)	(0.0326)	(0.0314)	(0.0351)
Urban	-0.0108	-0.0168	0.0228	0.000388
	(0.0737)	(0.0716)	(0.0673)	(0.0715)
Java	0.0230	0.0248	0.0797	0.145
	(0.339)	(0.330)	(0.286)	(0.298)
Log per capita oil cons.		0.0607**		0.0814***
		(0.0258)		(0.0256)
Log per capita sugar cons.		-0.0196		0.0829***
		(0.0311)		(0.0276)
Constant	11.90***	11.52***	11.40***	9.642***
	(0.422)	(0.669)	(0.541)	(0.868)
Observations	68882	68882	30078	30078
chi2	182851.5	193464.7	204887.9	184330.9

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Endogenous variable: the number of daily cigarettes smoked by current smokers
Instrumental variable: community's average daily cigarettes smoked by current smokers

F.2 Second-stage regression of the number of daily cigarettes smoked by current smokers on BMI: current-smoker and former-smoker (CSFS) Sample

Second-stage regression of the number of daily cigarettes smoked by current smokers on BMI: CSFS sample

	All (1)	All (2)	Male (1)	Male (2)
The number of daily cigarettes smoked by current smokers	-0.0310	-0.0924	-0.0562	-0.113
	(0.109)	(0.133)	(0.107)	(0.131)
Age	0.315***	0.353***	0.330***	0.366***
	(0.0683)	(0.0832)	(0.0679)	(0.0823)
Age square	-0.00256***	-0.00301***	-0.00276***	-0.00319***
	(0.000808)	(0.000985)	(0.000814)	(0.000990)
Married	0.0833	0.0632	0.0829	0.0668
	(0.0838)	(0.0907)	(0.0847)	(0.0916)
Hours of work	0.00147	0.00284	0.00182	0.00313
	(0.00240)	(0.00293)	(0.00248)	(0.00301)
Years of schooling	0.0522**	0.0649**	0.0634**	0.0762**
	(0.0264)	(0.0314)	(0.0282)	(0.0336)
Log effective per capita income	0.0753*	0.0932**	0.0876**	0.106**
	(0.0400)	(0.0470)	(0.0434)	(0.0512)
Urban	0.0422	0.0106	0.0487	0.0149
	(0.0708)	(0.0762)	(0.0753)	(0.0828)
Java	0.0622	0.140	0.0870	0.172
	(0.317)	(0.337)	(0.329)	(0.350)
Log per capita oil cons.		0.0669*		0.0681*
		(0.0356)		(0.0391)
Log per capita sugar cons.		0.0889***		0.0849***
		(0.0309)		(0.0283)
Constant	12.53***	10.60***	12.24***	10.35***
	(0.904)	(1.547)	(0.940)	(1.568)
Observations	26728	26728	24816	24816
chi2	179762.7	162985.6	172997.5	153653.8

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Endogenous variable: the number of daily cigarettes smoked by current smokers
Instrumental variable: community's average daily cigarettes smoked by current smokers

F.3 Second-stage regression of the number of daily cigarettes smoked by current smokers on BMI: current-smoker (CS) sample

Second-stage regression of the number of daily cigarettes smoked by current smokers on BMI: CS sample

	All (1)	All (2)	Male (1)	Male (2)
The number of daily cigarettes smoked by current smokers	-0.0150	-0.0443	-0.0243	-0.0534
	(0.0571)	(0.0634)	(0.0581)	(0.0646)
Age	0.299***	0.314***	0.306***	0.321***
	(0.0299)	(0.0330)	(0.0305)	(0.0337)
Age square	-0.00246***	-0.00263***	-0.00256***	-0.00272***
	(0.000341)	(0.000376)	(0.000351)	(0.000388)
Married	0.0210	0.0152	0.0452	0.0413
	(0.0809)	(0.0822)	(0.0813)	(0.0828)
Hours of work	0.000477	0.000925	0.000251	0.000704
	(0.00104)	(0.00112)	(0.00109)	(0.00118)
Years of schooling	0.0324**	0.0354**	0.0354**	0.0386**
	(0.0140)	(0.0146)	(0.0148)	(0.0154)
Log effective per capita income	0.0567**	0.0649**	0.0580**	0.0669**
	(0.0276)	(0.0292)	(0.0294)	(0.0313)
Urban	0.00855	-0.00809	0.0224	0.00523
	(0.0766)	(0.0778)	(0.0782)	(0.0797)
Java	0.438	0.528	0.498	0.594*
	(0.336)	(0.348)	(0.343)	(0.357)
Log per capita oil cons.		0.0588**		0.0634**
		(0.0253)		(0.0263)
Log per capita sugar cons.		0.0749***		0.0689**
		(0.0280)		(0.0277)
Constant	13.02***	11.67***	12.92***	11.58***
	(0.399)	(0.649)	(0.418)	(0.674)
Observations	23047	23047	21642	21642
chi2	166821.6	162209.3	165467.3	159579.7

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Endogenous variable: the number of daily cigarettes smoked by current smokers
Instrumental variable: community's average daily cigarettes smoked by current smokers

Appendix G. Estimations of the impacts of former-smoking behaviours on BMI by using fixed effects (FE)

G.1 Regression results of the effects of former-smoking status on BMI

Regression results of the effects of former-smoking status on BMI

	All	Male	Female
Former smoking status	-0.193**	-0.109	-0.388
	(0.0853)	(0.0880)	(0.275)
Age	0.479***	0.414***	0.488***
	(0.00893)	(0.0186)	(0.0101)
Age Square	-0.00380***	-0.00329***	-0.00388***
	(0.0000994)	(0.000203)	(0.000114)
Married	0.284***	0.304**	0.299***
	(0.0556)	(0.127)	(0.0617)
Urban	-0.0138	-0.00120	-0.0211
	(0.0578)	(0.112)	(0.0662)
Java	-0.644***	-0.449	-0.660**
	(0.234)	(0.367)	(0.278)
Log effective per capita income	0.0869***	0.126***	0.0813***
	(0.0157)	(0.0323)	(0.0176)
Log per capita oil cons.	0.0993***	0.0790**	0.105***
	(0.0193)	(0.0363)	(0.0223)
Log per capita sugar cons.	0.0275	0.0895**	0.0141
	(0.0195)	(0.0385)	(0.0224)
Hours of work	0.00115*	0.00289**	0.000825
	(0.000591)	(0.00117)	(0.000669)
Years of schooling	-0.0109	0.0727***	-0.0305**
	(0.0111)	(0.0194)	(0.0133)
Constant	9.432***	8.543***	9.666***
	(0.305)	(0.579)	(0.349)
Observations	49533	11626	37610

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.2 Regression results of the effects of the number of daily cigarettes smoked by former smokers before giving up smoking on BMI

Regression results of the effects of the number of daily cigarettes smoked by former smokers before giving up smoking on BMI

	All	Male	Female
The number of daily cigarettes smoked by former smokers before giving up smoking	-0.00469	-0.00199	-0.00622
	(0.00564)	(0.00562)	(0.0275)
Age	0.480***	0.417***	0.488***
	(0.00902)	(0.0194)	(0.0101)
Age Square	-0.00382***	-0.00335***	-0.00388***
	(0.000101)	(0.000214)	(0.000114)
Married	0.282***	0.324**	0.296***
	(0.0559)	(0.130)	(0.0618)
Urban	-0.0135	0.0155	-0.0226
	(0.0583)	(0.115)	(0.0663)
Java	-0.633***	-0.372	-0.660**
	(0.237)	(0.381)	(0.279)
Log effective per capita income	0.0856***	0.120***	0.0809***
	(0.0159)	(0.0335)	(0.0177)
Log per capita oil cons.	0.103***	0.102***	0.103***
	(0.0195)	(0.0376)	(0.0224)
Log per capita sugar cons.	0.0244	0.0753*	0.0146
	(0.0197)	(0.0400)	(0.0225)
Hours of work	0.00106*	0.00276**	0.000757
	(0.000596)	(0.00121)	(0.000671)
Years of schooling	-0.0124	0.0721***	-0.0309**
	(0.0112)	(0.0199)	(0.0133)
Constant	9.446***	8.440***	9.692***
	(0.307)	(0.600)	(0.350)
Observations	48054	10405	37373

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.3 Regression results of the effects of duration of smoking of former smokers before giving up smoking on BMI

Regression results of the effects of duration of smoking of former smokers before giving up smoking on BMI

	All	Male	Female
Duration of smoking of former smokers before giving up smoking	-0.00755*	-0.00508	-0.00855
	(0.00440)	(0.00467)	(0.0127)
Age	0.480***	0.418***	0.489***
	(0.00904)	(0.0197)	(0.0101)
Age Square	-0.00382***	-0.00335***	-0.00388***
	(0.000101)	(0.000218)	(0.000114)
Married	0.279***	0.312**	0.295***
	(0.0560)	(0.131)	(0.0618)
Urban	-0.0172	-0.000387	-0.0227
	(0.0583)	(0.114)	(0.0663)
Java	-0.648***	-0.456	-0.659**
	(0.235)	(0.362)	(0.279)
Log effective per capita income	0.0854***	0.118***	0.0810***
	(0.0159)	(0.0337)	(0.0177)
Log per capita oil cons.	0.102***	0.100***	0.104***
	(0.0195)	(0.0375)	(0.0224)
Log per capita sugar cons.	0.0229	0.0708*	0.0137
	(0.0198)	(0.0401)	(0.0225)
Hours of work	0.00106*	0.00277**	0.000752
	(0.000596)	(0.00121)	(0.000671)
Years of schooling	-0.0132	0.0707***	-0.0313**
	(0.0113)	(0.0200)	(0.0133)
Constant	9.481***	8.600***	9.694***
	(0.308)	(0.601)	(0.350)
Observations	47979	10345	37362

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

G.4 Regression results of the effects of length of time since quitting smoking on BMI

Regression results of the effects of length of time since quitting smoking on BMI

	All	Male	Female
Length of time since quitting smoking	-0.00727*	-0.00301	-0.0205
	(0.00413)	(0.00436)	(0.0155)
Age	0.480***	0.417***	0.488***
	(0.00901)	(0.0193)	(0.0101)
Age Square	-0.00381***	-0.00334***	-0.00387***
	(0.000101)	(0.000214)	(0.000114)
Married	0.282***	0.321**	0.296***
	(0.0559)	(0.130)	(0.0618)
Urban	-0.0134	0.0164	-0.0224
	(0.0583)	(0.115)	(0.0663)
Java	-0.649***	-0.451	-0.660**
	(0.235)	(0.362)	(0.279)
Log effective per capita income	0.0860***	0.122***	0.0812***
	(0.0159)	(0.0335)	(0.0177)
Log per capita oil cons.	0.102***	0.101***	0.103***
	(0.0195)	(0.0375)	(0.0224)
Log per capita sugar cons.	0.0237	0.0728*	0.0142
	(0.0197)	(0.0400)	(0.0225)
Hours of work	0.00104*	0.00266**	0.000764
	(0.000596)	(0.00121)	(0.000671)
Years of schooling	-0.0122	0.0725***	-0.0309**
	(0.0112)	(0.0198)	(0.0133)
Constant	9.462***	8.497***	9.693***
	(0.307)	(0.597)	(0.350)
Observations	48048	10403	37371

Cluster-robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix H. Sensitivity analyses

H.1 Robustness checks: balanced panel

H.1.1. A screen capture of the strongly-balanced data

```
xtset pidlink_num year
      panel variable:  pidlink_num (strongly balanced)
       time variable:  year, 1997 to 2014, but with gaps
               delta:  1 unit
```

H.1.2. Screen captures of the number of observations before any regressions

wave	Freq.	Percent	Cum.
1	5,332	20.00	20.00
2	5,332	20.00	40.00
3	5,332	20.00	60.00
4	5,332	20.00	80.00
5	5,332	20.00	100.00
Total	26,660	100.00	

Male	Freq.	Percent	Cum.
0	15,945	61.01	61.01
1	10,190	38.99	100.00
Total	26,135	100.00	

H.1.3. *Robustness check: estimation results of regressions with a strongly-balanced panel data*

Robustness check: balanced sample						
	Full Sample		**Current-smoker and non-smoker (CSNS) Sample**		**Current-smoker and former-smoker (CSNS) Sample**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Current smoking status	20.36	-2.34	13.04	-8.867	9.227	0.945
	(17.08)	(39.83)	(8.045)	(49.98)	(7.866)	(2.397)
Former smoking status	15.59	-1.384				
	(12.73)	(24.62)				
Observations	6167	9310	5529	9170	5116	372
	Former-smoker and non-smoker (FSNS) Sample		**Current-smoker and non-smoker (CSNS) Sample**		**Current-smoker and former-smoker (CSFS) Sample**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Daily cigarettes smoked by current smokers			-0.162	2.826	-0.00343	0.0827
			(0.133)	(2.877)	(0.281)	(0.671)
Former smoking status	-0.083	0.343				
	(0.180)	(0.360)				
Observations	1492	8986				
The number of cigarettes consumed per day before quitting smoking	-0.0167	0.0269**				
	(0.0140)	(0.0128)				
Observations	1492	8986				
Duration of smoking before quitting smoking	-0.00273	-0.00273				
	(0.00820)	(0.00820)				
Observations	1492	1492				
Length of time since quitting smoking	0.00304	-0.0111				
	(0.00915)	(0.0152)				
Observations	1492	8986	5476	9165	5062	367

Cluster-robust standard errors in parentheses

*p<0.1, **p<0.05, ***p<0.01

Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption.

Instrumental variables: percentage of current smokers in the community for current smoking status and community's average daily number of cigarettes consumed for smoking intensity.

H.2 Robustness checks: estimations by age groups

Robustness check: estimations by age groups						
	Current-smoker and non-smoker (CSNS) Sample					
	Male			Female		
	15-40 years	**41-60 years**	**>60 years**	**15-40 years**	**41-60 years**	**>60 years**
Current smoking status	5.792	8.693	-7.761	-97.50*	20.71	-3.803
	(5.174)	(20.20)	(5.104)	(58.67)	(54.23)	(13.27)
Observations	19595	8588	2205	24919	10428	2663
	Current-smoker and non-smoker (CSNS) Sample					
	Male			Female		
	15-40 years	**41-60 years**	**>60 years**	**15-40 years**	**41-60 years**	**>60 years**
Daily cigarettes smoked by current smokers	-0.160*	0.0974	0.271	-102.5	5.433	1.629
	(0.0970)	(0.145)	(0.490)	(334.9)	(8.228)	(2.212)
Observations	19430	8482	2166	24914	10417	2661
	Former-smoker and non-smoker (FSNS) Sample					
	Male			Female		
	15-40 years	**41-60 years**	**>60 years**	**15-40 years**	**41-60 years**	**>60 years**
Former smoking status	-0.0367	-0.0436	0.175	-0.514	-0.286	-0.793***
	(0.164)	(0.153)	(0.272)	(0.527)	(0.477)	(0.266)
Observations	7650	2947	1029	24847	10169	2594

Cluster-robust standard errors in parentheses
*p<0.1, **p<0.05, ***p<0.01
Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption.
Instrumental variables: percentage of current smokers in the community for current smoking status and community's average daily number of cigarettes consumed for smoking intensity.

H.3 Robustness checks: female elderly (above 60 years old)

Robustness check: the effects of former smoking among female elderly				
	Former-smoker and non-smoker (FSNS) sample			
	Female aged above 60 years			
	(1)	**(2)**	**(3)**	**(4)**
Former smoking status	-0.793***			
	(0.266)			
The number of cigarettes consumed per day before quitting smoking		-0.0485*		
		(0.0266)		
Duration of smoking before quitting smoking			-0.0173**	
			(0.00818)	
Length of time since quitting smoking				-0.0670***
				(0.0144)
Observations	2594	2549	2543	2547

Cluster-robust standard errors in parentheses
*p<0.1, **p<0.05, ***p<0.01
Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption.

H.4 Robustness checks: estimations for people with overweight/obesity status

Robustness check: estimations for people with overweight/obesity status				
	Current-smoker and non-smoker (CSNS) sample			
	At least once pre-obese or obese		**Pre-obese or obese**	
	Male	**Female**	**Male**	**Female**
Current smoking status	4.264	-95.21	1.416	0.0779
	(3.972)	(103.4)	(5.772)	(22.50)
Observations	12546	23837	8604	17066
	Current-smoker and non-smoker (CSNS) sample			
	At least once pre-obese or obese		**Pre-obese or Obese**	
	Male	**Female**	**Male**	**Female**
Daily cigarettes smoked by current smokers	-0.202	14.48	-0.14	8.349
	(0.222)	(11.29)	(0.339)	(9.394)
Observations	12411	23825	8502	17055
	Former-smoker and Non-smoker (FSNS) sample			
	At least once pre-obese or obese		**Pre-obese or obese**	
	Male	**Female**	**Male**	**Female**
Former smoking status	-0.0312	-0.354	0.0596	-0.0560
	(0.118)	(0.371)	(0.143)	(0.456)
Observations	5821	23617	4080	16884

Cluster-robust standard errors in parentheses

*p<0.1, **p<0.05, ***p<0.01

Control variables: age, age square, marital status, total working hours per week, years of schooling, log effective per capita income, dummy of urban location, dummy of java island, log effective per capita oil consumption, and log effective per capita sugar consumption.

Instrumental variables: percentage of current smokers in the community for current smoking status and community's average daily number of cigarettes consumed for smoking intensity.